

Don't mention the war...

The United States could taint the ITER project to build a prototype magnetic-confinement fusion reactor if it exploits the selection of a site to settle diplomatic scores. But its European partners don't need to let that happen.

Spencer Abraham, the US energy secretary, seems to have a preference over which European nation — France or Spain — should apply to host the proposed ITER fusion reactor. Not surprisingly, perhaps, given Spain's support for the Bush Administration's war on Iraq, and France's opposition, it is the Spanish bid that he has been talking up.

By the end of this year, the ITER partners are expected to make a decision on where to build the world's next magnetic-confinement fusion device, which would explore the behaviour of superheated hydrogen ions over long enough periods to infer how they might act in a fusion-based power plant. There are four proposed sites: one in Japan, two in Europe and one in Canada.

Most eyes are turned to the two parties whose money has sustained the project so far: Japan and the European Union (EU). Russia is a long-standing partner, but can't afford to offer a site, and certainly isn't planning to advise Europe and Japan on its preference. Canada has offered a site in Ontario, but although its selection would suit participating researchers in the neighbouring United States, US officials have never spoken out in favour of the Canadian site. They must realize that such a public statement of support, from a nation that has only just decided to rejoin the project and is planning to contribute no more than 10% of its US\$5-billion costs, would be inappropriate.

This helps to explain why Abraham's comments, made on a visit

to Madrid earlier this month, are causing some disquiet. He came perilously close to expressing US support for the Spanish bid to host ITER (see page 211). This is liable to offend the French government, as Abraham well knows.

France is one of the strongest European supporters of magnetic-confinement fusion research in general, and of ITER in particular. Its Atomic Energy Commission has the technical and managerial expertise to take a leading role in the project, so the French bid is a very serious contender. Spain, meanwhile, also has a solid bid, and can argue to that its relatively young democracy, and less-well-developed physics community, means that it has more to gain from hosting such a prestigious international project.

Abraham could remove the sour taste left by his comments in Madrid by stating unequivocally that the choice of a European site for ITER is a matter for the EU. This would prevent the project being sucked into the vortex of recrimination created by the current US administration's self-avowed intention to punish France for having the gall to oppose its invasion of Iraq.

But even if Abraham declines to make such a statement, the EU can defuse the issue. Earlier this week, ministers from its member states agreed that the EU must soon come down in favour of just one of the European sites. If the project's European partners can make a clear choice between its two bidders before the final site selection is made, Abraham's untimely intervention will be moot. ■

Gene patents and the public good

A race to claim patents on the SARS virus raises questions about the patent system's ability to cope with genomics.

Barely had the last nucleotide in the genetic code of the SARS coronavirus been read when the race to claim intellectual rights to the sequence began. And among those filing was the US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, in what was essentially a pre-emptive strike (see page 214). CDC director Julie Gerberding told reporters last week that, in private hands, a patent on the viral sequence might delay the development and refinement of tests and treatments for the contagious pneumonia that has already killed several hundred people.

Gerberding's admission gives tacit acknowledgement to a growing concern among biomedical researchers that broad patents on genetic sequences may, in some cases, have a stifling effect on research and negative consequences for public health. Take the struggle over predictive testing for breast cancer based on the *BRCA1* and *BRCA2* genes. Since the Utah-based company Myriad Genomics won a European patent on the genes in 2001, the Curie Institute in Paris, with the backing of the French government, has been fighting for the right to continue testing women for the genes, which it can do for about a third of what Myriad charges. A similar situation exists in Canada, where several provincial governments have told their health authorities to continue testing in the hope of ultimately winning in court the right to do so.

Although the extent of the problem is hard to quantify, the calls

for action are growing louder. Britain's Nuffield Council on Bioethics last year recommended that in future, patents on genetic sequences should be the exception rather than the rule. In Canada, a report issued by Ontario's provincial government, following its experience with *BRCA1* and *BRCA2*, explored solutions such as a compulsory licensing scheme in which a public entity would determine who should have access to which gene patents, and even set licensing fees. The Canadian Biotechnology Advisory Committee is now set to move the proposal on to the ministerial level in the federal government.

In the United States, however, there currently seems to be little momentum for change. Last year, US Representative Lynn Rivers (Democrat, Michigan) sponsored a bill in Congress that would have allowed healthcare entities to perform genetic diagnostic tests without fear of patent-infringement lawsuits. It would also have guaranteed an exemption from patent restrictions for basic research. But Rivers has since lost her seat in the House and the legislation has little hope of revival.

There are no clear-cut answers. But when pre-emptive patenting is necessary to ensure that rapid solutions are found to an important health problem, something seems to be out of balance. Policy-makers should investigate what checks and balances are necessary to ensure that the patent system continues to do its job of stimulating innovation for the public good. ■

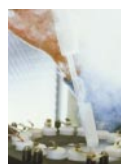




Off beam
Cloudy horizon
for Australian
synchrotron
p210



All at sea
Mixed outlook for
battered cod stocks
across the Atlantic
p212



Cool response
Most frozen embryos
can't be used for
stem-cell research
p213



In the red
Giant jellyfish
makes a splash in
the species stakes
p214

Biologists take tentative steps towards bespoke cancer drugs

David Cyranoski, Tokyo

Japanese biologists are vying with each other to identify patients who are likely to develop side-effects from a highly promising cancer therapy.

Their pioneering efforts to screen for side-effect susceptibility will help to shape future drug development and clinical trials, as expectations grow for prescriptions that are better tailored to an individual's genetic make-up or physiological condition.

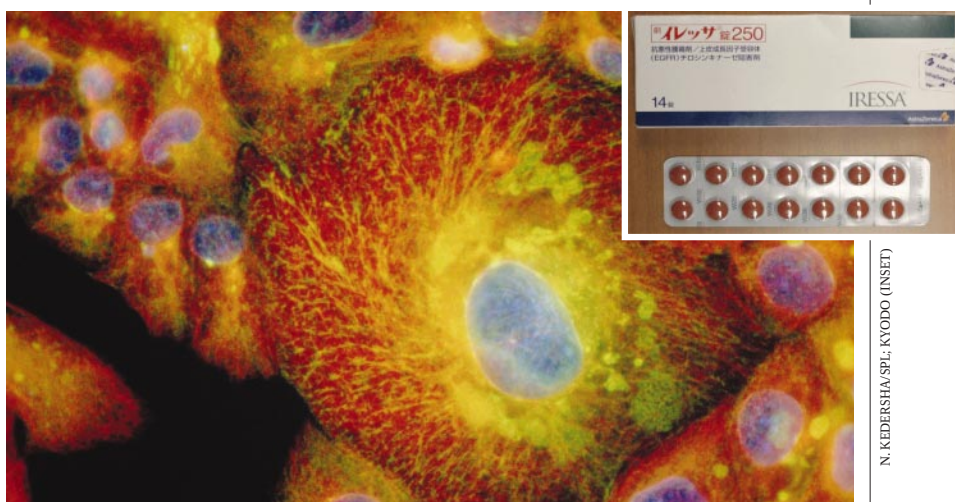
The studies are focused on the anticancer drug Iressa. Last July, Japan became the first country to approve the therapy, after it sped through the country's new fast-track drug-approval system. Despite the 246 deaths so far ascribed to its side-effects, 28,000 Japanese lung-cancer patients are using Iressa, and the United States and Australia this month approved the drug.

Produced by Anglo-Swedish company AstraZeneca, Iressa is one of a revolutionary new generation of therapies based on an understanding of the specific gene mutations that cause individual cancers (see *Nature* **416**, 470–474; 2002). The drug is a tyrosine-kinase inhibitor, which blocks specific growth-factor receptors — proteins on the cell membrane that pass on signals commanding the cell to grow and are overactivated in cancer.

Despite hopes that this type of drug would cause side-effects in very few patients, 616 people receiving Iressa in Japan have developed severe reactions — mostly a pneumonia-like illness called interstitial lung disease (ILD). Nonetheless, health officials say that the drug works too well for it to be recalled given the lack of other options for treating the disease. The health ministry says that various hospitals throughout the country have reported reductions in tumour size in between one-sixth and one-third of cases.

The two studies, which will begin shortly, are collaborations between universities and AstraZeneca. One will track the levels of certain blood proteins in different patients, and the other will examine the genetic differences between individuals called single-nucleotide polymorphisms (SNPs).

The SNPs project, led by Yusuke Nakamura of the University of Tokyo, will scour some



On target: Iressa (inset) inhibits proteins that control the growth of lung cancer cells.

4,000 SNPs in 200 genes to look for patterns that correlate with a tendency to experience side-effects. If they cannot find a pattern in this limited study, they will expand the range genome-wide to some 200,000 SNPs.

The second study will be led by Yoshiji Fujita, head of Tokyo Medical University's new Clinical Proteome Center. Fujita's team will take blood samples from 500 patients before and after they first take Iressa, and again after two months, by which time any side-effects should have developed. The group will look for any change in protein levels that correlates with later development of ILD. If all goes well, Fujita hopes that by next summer he will have a protein chip under development that will be able to tell patients within 12 hours of them starting the therapy whether they should stop taking Iressa or adjust its dosage.

But questions remain about both approaches. SRL in Tokyo, Japan's largest clinical-testing company, recently decided to devote its entire science programme to gene and protein expression, foregoing SNPs. "There are just too many SNPs, and finding the target is too difficult," says chief technology officer Kazumasa Hikiji. "Expression studies are more accurate and more cost-effective."

And sceptics argue that SNPs may not

provide a definitive marker when side-effects turn out to be caused by non-genetic factors, such as the intake of other drugs. Even Nakamura accepts that there is no guarantee it will work. "If many genes are involved, and each has a very limited effect, it may be difficult to get a positive result," he concedes.

The protein studies "make good sense", according to Charles Sawyers of the Jonsson Comprehensive Cancer Center at the University of California, Los Angeles. However, "the question is whether the pattern is strong enough to be seen", he says. "I think there is a risk that both approaches could fail, as the clinical application here is new. But the scientific rationale behind both is quite strong."

For AstraZeneca, any news about what is going on will be good news. "Ideally the tests will say: 'the genetic factors are like this and the protein expression is like this', and in that gap, we will be able to figure out how Iressa is involved," says Michio Tanaka, team leader for the company's oncology and therapy research division in Osaka.

The repercussions could be far-reaching, according to Fujita. "If we succeed, patients and drug-approval agencies will begin to expect studies on early detection of side-effects from pharmaceutical companies that are doing clinical trials," he says. ■

N. KEDERSHA/SPL; KYODO (INSET)

Researchers rue hasty destruction of wheat

Carina Dennis, Sydney

Hindsight is always 20/20, as researchers at one of Australia's top plant-research labs have just discovered. Last month, the laboratory destroyed most of its wheat samples — more than 40,000 plants in all — because of fears that they were housing wheat streak mosaic virus, a disease that farmers desperately want to keep out of Australia.

But now it looks as if the virus is already established in Australia anyway, leaving laboratory managers worried that they might have wiped out up to a year's worth of work for nothing.

The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Australia's largest research agency, first identified the virus — which can badly affect crop productivity — at two of its breeding stations in Canberra. The stations are part of CSIRO Plant Industry, an agricultural laboratory run by the agency.

"We originally thought the virus was just at our Canberra sites and we thought we could eradicate it. We didn't know that it was already in the country," says T. J. Higgins, assistant director of the lab. "Faced with the same decision, I would do it again."

The virus has since been found at several more research facilities. But these samples have been quarantined, instead of being

destroyed, as part of a government-instigated assessment of the spread of the virus in Australia's research facilities and wheatfields.

"I feel confident that the virus has been here for a long time — at least since the 1990s," says Anthony Rathjen, a wheat breeder at the University of Adelaide's Waite campus, the second site at which the virus was detected.

But the conclusion that the virus is already found in Australia comes a bit late for the CSIRO plant laboratories. "We're currently estimating the costs and it could take a year or so for some breeding programmes to recover," says Higgins. "Morale was initially very low, but people are now bouncing back."

Wheat producers are not very happy either. "We depend on new varieties, and the destruction at the CSIRO has meant that a lot of the valuable material has been lost," says Murray Jones, senior vice-president of the Canberra-based Grains Council of Australia. Nonetheless, he adds, "at the time, the right decision was made."

Farmers remain hopeful that, in Australia, the virus is confined to plant-breeding laboratories, where the vector mite that transmits it enjoys a constant supply of young, green wheat. "I think it's a great nuisance for glasshouses but is insignificant as far as the industry is concerned," says Rathjen.



Binned: researchers killed 40,000 wheat plants in a bid to quell wheat streak mosaic virus.

The virus is already endemic in the wheatfields of the United States, where it results in a small but significant loss of productivity. But Australian wheat farmers want it kept out. "Every continent is different and you can't extrapolate," says Higgins. "The disease might be much worse or much less damaging for Australia."

Cash crisis casts pall over synchrotron X-ray source

Carina Dennis, Sydney

Plans to build a synchrotron X-ray source in Melbourne could be heading for trouble, researchers say, after federal and state government officials clashed over who should foot the bill.

The state government of Victoria has pledged the A\$157 million (US\$102 million) needed to build the facility, the construction

of which should begin later this year. But project planners need another A\$50 million to build the 'beamlines', which will channel the X-rays produced as particles shoot around the synchrotron's central ring.

Victoria had hoped that the federal government would help to pay for the beamlines. "We have punched well beyond our weight, and now it's time for the federal government to come on board," says John Brumby, Victoria's innovation minister.

But on 11 May, the federal science minister, Peter McGauran, said that the government would not allocate money specifically for the synchrotron. It will instead have to compete for funds from existing programmes, such as those of the Australian Research Council and the National Health and Medical Research Council. This means that any federal money for the project looks likely to come at the expense of scientists' grants.

The facility's prospective users fear that squabbling between the conservative, federal government and the Labour-led state government in Melbourne will damage

Australia's largest-ever scientific project.

"It is unthinkable that the federal government would not make some contribution to a facility of such great national benefit," says Peter Darvall, vice-chancellor of Monash University, where the synchrotron is to be built.

Researchers, who hope to use the synchrotron's X-rays to probe three-dimensional molecular structures, are exploring many avenues to find the money. "We envisaged that the scientific community would be principally responsible" for raising it, says Frank Larkins, chair of an advisory committee coordinating the construction of the beamlines and an administrator at the University of Melbourne, which has agreed to fund at least one beamline itself.

Larkins is still hopeful that the money can be found. "Once we have prepared the scientific case, we will be approaching a range of bodies, including the federal government, research agencies, universities and industry," he says. "If we don't have the commitment of funds in 18 months, then we will start to be concerned."

DESIGNING MELBOURNE

Synch or swim: one possible design of the synchrotron Australia plans to build.



Spencer Abraham (right) meets with Prime Minister José María Aznar during his visit to Spain.

US support for Spain triggers unease over fusion project

Geoff Brumfiel, Washington, and Declan Butler, Paris

The US energy secretary has come close to endorsing Spain's bid to host a huge international fusion experiment — snubbing French plans for the project and unsettling delicate European site negotiations.

Spencer Abraham visited Madrid on 1 May to discuss Spain's bid to host ITER, the planned US\$5-billion project to build an experimental magnetic-fusion reactor.

After meeting with high-level Spanish officials, Abraham held a press conference at which he described the proposed site at Vandellòs near Barcelona as “very impressive”. He emphasized that the trip was the first visit he had made with regard to ITER's siting, saying: “I hope that demonstrates the level of interest we have [in] the proposal Spain has offered,” (see www.embusa.es/ratoabraham.html).

He added that when President Bush brought the United States back into ITER earlier this year he was aware of Spain's proposal. “He asked me to travel to Madrid to learn more” about the site and “to make sure that the United States would give it very serious consideration”, Abraham said.

Abraham's visit — his first to Spain — preceded a meeting of the European Council of Ministers, which agreed to select a single European site for the reactor this year.

Spanish and American officials deny that the United States is meddling in Europe's affairs. But European researchers say that they detect an attempt by the United States to reward Spain for its cooperation in the occupation of Iraq, which France opposed.

ITER is an international collaboration between the European Union, Canada, Japan, Russia, China and the United States. At present, there are four proposed sites for the facility: Clarington in Canada, Rokkasho in Japan, Cadarache in France and Vandellòs.

Jeanne Lopatto, a spokeswoman for the US energy department, denies that Abraham is playing favourites. “He did not by any means endorse one site over another,” she says. Last week, the Spanish economics minister, Rodrigo Rato, similarly dismissed as “simplistic” the idea that Spain is being rewarded for its cooperation on Iraq.

Christian Poncet, who is coordinating the French bid to host ITER, plays down Abraham's comments. “I think that Mr Abraham will also visit Clarington, Rokkasho, and I hope he will visit Cadarache,” he says.

And even if the United States were to express a preference, that need not be perceived as meddling in European affairs, suggests Murray Stewart, who heads ITER Canada. The preferences of other ITER partners are “reasonable criteria” for choosing a domestic site, Stewart says.

An official at the European Commission disagrees, saying that the selection of a European site is an internal matter. Abraham's remarks are “at the very least surprising”, he adds.

Other officials in Europe are nervous over what they see as the US administration's agenda infringing on their decision-making process. “The indications we've had from the State Department is that the United States would stay neutral,” says one German official. Another senior European fusion researcher, who didn't want to be named, says that many in his community take a “dim view” of the energy secretary's actions.

The European Commission is now undertaking a technical assessment of the two sites, probably to be led by David King, chief scientific adviser to British Prime Minister Tony Blair. King visited Spain on 9 May and told reporters he found Vandellòs “very attractive”. But he added: “Right now, I'm not sure which site is better.”

Medical council pins hopes on public advocate of science

Jim Giles, London

Leading neuroscientist Colin Blakemore has been chosen to run Britain's main biomedical research agency, the Medical Research Council (MRC).

Blakemore's assertive personality, which colleagues say he brings to all aspects of his work, is likely to be tested to the full at the London-based council. The organization that he inherits from George Radda, the molecular cardiologist who has been the MRC's chief executive since 1996, has been facing a barrage of criticism lately from the community it funds (see *Nature* 422, 461; 2003).

Currently head of the Centre for Cognitive Neuroscience at the University of Oxford, Blakemore is well respected for his research on the development of the visual system. He also has a strong track record of promoting science with the general public.

Blakemore, who takes over from Radda in October, says that he will investigate whether the MRC grant applications process could be made more open and its paperwork streamlined. He also says he would like to see whether the council can help to increase the credit that researchers get for taking part in science communication. “Young scientists need to get more professional reward for these activities,” he says.

Blakemore is often in the media spotlight. During the mid-1990s he was singled out by animal-rights campaigners and received numerous death threats. A letter bomb was sent to his house, although the device did not explode. The experiences prompted Blakemore to become an unofficial spokesman for the use of animals in research — a role he says he will continue to play after he takes up his new position.

Colleagues of Blakemore expect him to tackle this and other issues with his



Colin Blakemore: energetic science communicator.

usual, and sometimes fiery, dynamism. “He can get quite heated,” notes Andrew Lumsden, a developmental neurobiologist at King's College London. But Lumsden, like other researchers contacted by *Nature*, says that such passion will work in the council's favour.

Europe dithers as Canada cuts cod fishing

Quirin Schiermeier, Munich

Cod on opposite sides of the Atlantic Ocean face contrasting prospects as authorities in North America and Europe try to conserve their once-plentiful stocks.

Late last month, the Canadian fisheries ministry ordered an end to all fishing of Atlantic cod (*Gadus morhua*) in three regions off the coast of Newfoundland and Labrador. It also set up a Can\$50-million (US\$36-million) action plan to assist fishermen and communities that will be affected.

But such closures are not yet envisaged in Europe — although fisheries experts are calling for a similar moratorium. Last week, the European Commission outlined a ten-year recovery plan for cod, which calls for lower catch quotas and a better monitoring system.

Although cod stocks in the North and Irish Seas have declined by 90% since the 1960s, some Canadian populations are down by more than 99%. Cod in warmer European waters reach maturity at an earlier age — an advantage in terms of reproductive ability.

On 2 May, the independent Committee on the Status of Endangered Wildlife in Canada (COSEWIC) recognized Atlantic cod as an endangered species after finding that some stocks face imminent extinction.

"If Canadian cod is threatened by extinction, this should ring alarm bells in Europe," says Georgina Mace, scientific director of the Institute of Zoology in London, and a member of a committee that identifies endangered species at the World Conservation Union.



Cold comfort: fishing may be reduced in some areas, but Atlantic cod could still face extinction.

"Harvesting large long-lived species on a sustainable basis is much more difficult than we used to believe. Populations don't seem to bounce back just because fishing effort is being reduced for a while," Mace says.

COSEWIC member Jeff Hutchings, a marine biologist at Dalhousie University in Halifax, Nova Scotia, has shown that depleted fish populations do not recover easily when fishing pressure is reduced or stopped (see *Nature* **419**, 662–665; 2002). "This is another reason not to let populations drop below a critical size," he says. "Fisheries

managers in Europe should take note."

But limiting fisheries without offering fishermen a viable alternative is unrealistic, argues Colin Bannister, of the UK government's Centre for Environment, Fisheries and Aquaculture Science in Suffolk. "As much is being done as can be done, bearing in mind the socio-economic situation," he says.

In Europe, some half-a-million people make a living from fisheries, including ship-building and repair. Attempts to reduce fleets or quotas have met fierce opposition.

Last December, thousands of British fishermen took to the streets in an unsuccessful protest against tighter European Union (EU) quotas on cod fishing. These were brought in on the basis of advice from the Copenhagen-based International Council for the Exploration of the Sea (see *Nature* **419**, 866; 2002).

Scientists, however, hope that the long-planned fisheries reform will maintain its momentum, and that scientific advice will gain more influence on European policy.

"Europe has got to get out of the old system where lobbying of the fisheries industry and disagreement between EU member states have each year watered down the best available scientific advice," says John Reynolds, a marine ecologist at the University of East Anglia in Norwich.

Bannister welcomes the European Commission's plan to set up regional advisory bodies that will incorporate different interest groups, including fishermen and scientists.

In Canada such efforts are well under way already. Next month, scientists, fishermen and industry representatives at the third North Atlantic Responsible Fishing Conference in Yarmouth, Nova Scotia, will discuss the scientific, social and economic issues related to fishing.

Asteroid probe to test technologies

David Cyranoski, Tokyo

A compact spacecraft mission headed off last week on an ambitious return trip to retrieve tiny rock samples from an asteroid 300 million kilometres away.

The Japanese Muses-C mission is the first in which rocks are to be brought directly back from an asteroid. The mission is intended as a demonstration project for several critical technologies, including a solar-powered ion motor that will provide its main means of propulsion, and up-to-date imaging techniques to control it remotely when it lands. The craft is designed to collect a gram of rock from asteroid 1998SF36 and bring it back to Earth in 2007.

Project leader Jun'ichi Kawaguchi of the Institute of Space and Astronautical Science (ISAS) in Sagami-hara, near Tokyo, says the major goal of the mission is to convince researchers and government officials of the viability of a set of technologies for use in future

sample-return missions.

"There are a lot of bodies out there in space, and we want to show people that they can be studied like this," says Kawaguchi.

Muses-C is designed to pelt the surface of asteroid 1998SF36 with steel projectiles travelling at 300 metres per second and then gather some of the dust that they kick up. Kawaguchi says one gram will be enough for studies into the substance of the asteroid. ISAS and NASA researchers plan to study the sample for a year and then make them available to other interested scientists.

The substance of the asteroids remains a mystery to scientists, who have had to depend on remote sensing imaging, such as that undertaken by NASA's NEAR-Shoemaker mission (J. Veverka *et al.* *Nature* **413**, 390–393; 2001), for clues about their composition. ISAS built the 530-kg craft itself after NASA dropped plans in 2000 to send a robot with it. The mission will cost at least ¥12.7 billion (US\$108 million).

Low embryo count fuels US stem-cell debate

Kendall Powell, Denver

Only about 275 of the roughly 400,000 frozen embryos currently stored in the United States are likely to be of any use in stem-cell research, according to the first comprehensive inventory of the embryos.

The low number gives the lie to the notion that fertility clinics are awash with spare embryos that are suitable for deriving new lines of human embryonic stem cells, opponents of the research say. But its supporters counter that there are not many embryos because so few fertility clinics arrange for patients to donate embryos to research.

According to a survey by Jacob Mayer, an embryologist at the Jones Institute for Reproductive Medicine in Norfolk, Virginia, and colleagues, donors have approved only 11,000 of the embryos for use in research (D. I. Hoffman *et al. Fertil. Steril.* **79**, 1063–1069; 2003).

Many of the 11,000 embryos would not survive freezing and thawing, and only 25% of those that do are expected to develop to the blastocyst stage. Mayer and co-authors estimate that the process is so inefficient that just 275 would lead to new stem-cell lines.

"The American public had the impression that hundreds of thousands of embryos were out there that patients were eager to have destroyed in research," says Richard Doerflinger, a lobbyist at the US Conference of Catholic Bishops. "This shows that this

is clearly not the case."

"Fertility clinics may not contain the plethora of material they were expected to," says Mayer, whose institute conducted the survey of 340 fertility clinics in association with the Society for Assisted Reproductive Technology and RAND, a private non-profit research institution.

Researchers relying on frozen embryos from infertile couples are in for a shock, not only because of the scarcity but also because embryos from infertile couples may be less likely to develop normally, says Amy Sparks, director of the University of Iowa Reproductive Testing Laboratory in Iowa City. Her clinic, where about 300 embryos have been donated to research in the past 15 years, is one of just a handful in the country that has protocols for the use of donated embryos in stem-cell research.

Even a small number of suitable embryos is viewed by researchers as a potentially useful addition to the 11 stem-cell lines already sanctioned by the government for use in research. "That 275 is 20 times what we have now," says George Daley, a stem-cell researcher at the Whitehead Institute in Cambridge, Massachusetts. Even though thousands of embryos have been designated for research, he says, "there's no effective way of getting them into researchers' hands".

Some researchers predict that donations



Research on ice: less than 3% of frozen embryos in the United States could be used in the lab.

to research will rise as public awareness of stem-cell and other human embryo research increases. But opponents say that the numbers reflect the public's rejection of the use of embryos in research. "There's a widespread intuition in the public that embryos belong in uteruses," says Ben Mitchell of the non-profit Center for Bioethics and Human Dignity in Bannockburn, Illinois. ■

MIT pulls out of Asian Media Lab in argument over role

K. S. Jayaraman, New Delhi

An effort to replicate in India the world-famous Media Lab, run by the Massachusetts Institute of Technology (MIT), has collapsed amid angry recriminations between the Indian government and the original lab in Cambridge, Massachusetts.

Media Lab Asia, which opened in New Delhi amid much fanfare two years ago

(*Nature* **411**, 623; 2001), will continue without MIT's involvement after an acrimonious meeting last week between Arun Shourie, India's information-technology minister, and Nicholas Negroponte, chair of the MIT Media Lab's governing board.

MIT officials say that the decision to end the collaboration came after Shourie said he wanted practical technologies to flow from the laboratory, rather than the futuristic ideas and approaches for which the MIT lab won its reputation as one of the world's hottest research labs.

A Press Trust of India news agency report from New York quoted Negroponte as saying that MIT pulled out of the project because Shourie wanted to run it like other programmes in India. It said that Shourie does not believe in rural development through information technology, and that he is even less interested in basic innovation.

But Shourie told reporters that MIT's interpretation of events was "strange", and that India called off the contract as it "gained nothing" from the collaboration.

Media Lab Asia was MIT's second Media

Lab outside the United States following Media Lab Europe, which was established in Dublin, Ireland, in 2000. The Asian lab was set up in India after MIT considered offers to host it from several other countries, including China and Singapore. Its board had selected research projects to be carried out at five Indian Institutes of Technology (IITs) in Chennai, Mumbai, Kharagpur, New Delhi and Kanpur.

The initial agreement to run Media Lab Asia expired on 31 March and has not been renewed. Shourie says it failed not because he has no interest in rural development, but "because researchers in the five IITs were unable to quantify the contribution from MIT". Krithi Ramamritham, head of the Media Lab Asia's centre at the IIT in Mumbai, says that after the initial planning phase, the project got virtually no technical guidance from MIT.

Shourie adds that the project was failing to attract private funds, and that the Indian government didn't want to pay US\$5 million over ten years to use the name 'Media Lab', to which MIT claims the rights. ■



Discontent grows as cancer institute remains rudderless

Munich Germany's largest biomedical research institute, the Heidelberg-based German Cancer Research Centre (DKFZ), which has nearly 2,000 employees, has suddenly found itself without a leader — and with no immediate prospect of finding one.

Virologist Bernhard Fleckenstein was presented as successor to scientific director Harald zur Hausen last month, but a few days later he ended negotiations with the German federal research ministry. The ministry apparently failed to meet his desired salary, but Fleckenstein says that there were also “irreconcilable differences” between himself and the ministry about how the DKFZ should be run.

The DKFZ is one of Germany's 16 national research centres, which have recently been reoriented towards strategic research. Fleckenstein says he is uncomfortable with the hierarchical structure of decision-making in Germany's research programme, and wanted more freedom for DKFZ scientists.

The appointment committee has resumed what could be a lengthy search. In the meantime, scientists are worried about the power vacuum, given the threat of financial cuts as the German government tries to balance its deficits, and the impending retirement within the next few years of several of the institute's senior scientists.

Parapsychology post set up for long-dead benefactor

Stockholm After a 40-year wait, psychologists at Lund University in Sweden are about to establish a chair endowed by a legacy from Danish businessman Paul Thorsen. But the move raises even more eyebrows now than it would have done in 1962, as Thorsen stated that the new position should be in the areas

of his own amateur interests: parapsychology and hypnology.

Thorsen, a margarine magnate, had no family but wanted to ensure a pension for his two servants before allowing his legacy to be released to the university. Now that both have died, the university is advertising the position.

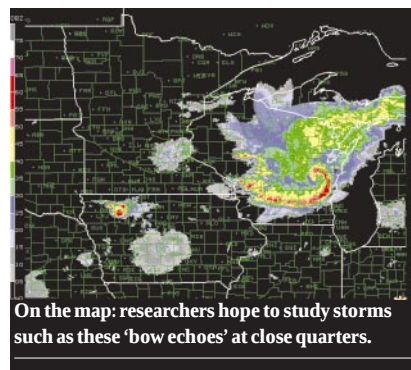
“We’ve got accustomed to the idea by now,” says Jitka Linden, head of the university's psychology department, “but we are aware that it might be controversial.” But she adds that “there are already several chairs in parapsychology around the world, and our selection will be as rigorous as for any other professorship”. If no suitable candidate can be found, she says, the chair will not be filled.

US acts to keep SARS sequence public

San Francisco The US Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, has applied for a US patent on the coronavirus thought to cause severe acute respiratory syndrome (SARS). The move is an attempt to head off companies that might monopolize future diagnostics and treatments for SARS, says CDC director Julie Gerberding. But the situation may develop into a full-scale intellectual-property battle.

Among the CDC's competitors for patent rights are the University of Hong Kong, where the coronavirus was first observed, and the British Columbia Cancer Agency, the private research foundation that first sequenced the virus's genome. The cancer agency says that it would plough any royalties back into research.

In joining the patent fray, the CDC, which seldom files such applications, is recognizing that patents on early-stage research can hinder further development, depending on how the holder behaves, says Robert Cook-Deegan, director of the Center for Genome Ethics, Law and Policy at Duke University in Durham, North Carolina. “They are smart to do it,” he says.



On the map: researchers hope to study storms such as these ‘bow echoes’ at close quarters.

Airborne lab crews hope study will go down a storm

Washington When this summer's thunderstorms begin barrelling across the US Midwest, researchers will be tagging along in the largest storm study since the 1970s.

Funded by the National Science Foundation, the \$4-million study will launch three aircraft, loaded with atmospheric scientists and mobile weather labs, from St Louis, Missouri, directly into a type of storm cluster called a mesoscale convective system.

Unlike thunderstorms, which can wax and wane within hours, these systems unfold in the evening and can devastate communities with tornado-like winds and rainfall for a full 24 hours. As they decay, they can trigger new storms called bow echoes that roll off the main storm, forming an ‘arc of torment’.

The flight data will help forecasters to understand what triggers these storms, and to predict where and when they will arise.

Ousted university vet sues over dismissal

San Diego A Chinese researcher who was acquitted last year of trying to steal tissue-growing technology from the University of California, Davis, is suing the university.

Bin Han, a veterinarian who worked in an ophthalmology lab at the university, claims in a state lawsuit filed on 5 May that allegations — which led to criminal charges — were made against him in retaliation for disclosing operational problems.

University officials deny this, arguing that Han's annual contract was simply not renewed last spring. Last May, police at the university seized research materials for corneal repair at Han's home, and found an air ticket to China, where he was collaborating in research.

Han alleges in the Sacramento Superior Court lawsuit that he was wrongfully dismissed, retaliated against as a whistleblower, and had his civil rights violated. He now wants a new job at the university. The lawsuit is expected to be closely watched by other US universities, given the sensitivity of investigations into foreign researchers.

Giant jellyfish leaves researchers red-faced

San Francisco How it managed to escape notice for so long is unclear, but a new giant species of jellyfish, measuring a metre across, has only now been described by a team of marine biologists in California. The monster jellyfish, known as ‘big red’, was first spotted in the inky waters at depths of more than 600 metres in 1993. But it was not until 1998 that George Matsumoto and his team at the Monterey Bay Aquarium Research Institute in Moss Landing first got close enough to realize that it could be new to science. Rather than having slender tentacles, big red (*Tiburonia granrojo*) has between three and seven chubby arms that it uses to grapple with prey, and has been assigned to an entirely new subfamily — the Tiburoniinae.



The man they love to hate

Bjørn Lomborg is reviled by green activists and has come under ferocious attack from many environmental scientists. Just why does he provoke such strong reactions, and how influential might his opinions become? Jim Giles investigates.



Meeting Bjørn Lomborg for the first time, it's hard to understand what all the fuss is about. For his upbeat assessment of the state of the world's environment, Lomborg has become the bogeyman of the green movement, has been accused of scientific misconduct, and has even been likened in the pages of *Nature* to those who deny the Holocaust.

Yet, in person, Lomborg is far from the ogre that this publicity might suggest. He is calm, friendly and utterly charming. He runs a new environmental research institute in Copenhagen that is disarmingly informal — as I arrived, a member of staff ambled down a corridor, brushing her teeth. Clad in jeans and a T-shirt, Lomborg seems to have more in common with the political liberals that he has so incensed than with the conservative establishment that has eagerly embraced his message.

Lomborg's notoriety stems from his 2001 book *The Skeptical Environmentalist*, a data-heavy assessment of the state of the planet that paints an extremely optimistic picture. The storm that the book generated has been well documented, but just why did the debate become so heated? Will the book, and Lomborg's continuing work, have any lasting influence? And what lessons does the Lomborg affair hold for those who want to promote informed public and political

debate about environmental science?

The roots of *The Skeptical Environmentalist* lie in the work of another man whom the greens love to hate: the late free-market economist Julian Simon of the University of Maryland at College Park. In 1997, Lomborg, then a lecturer applying statistics to problems in political science at the University of Aarhus in Denmark, came across a magazine article in which Simon rebutted many of the doomsday predictions that environmentalists have made about the planet. "When I read the article I thought 'hell, no,'" Lomborg recalls. "I thought that obviously the environment is getting worse. But Simon said one irritating thing: go check the facts."

Challenging task

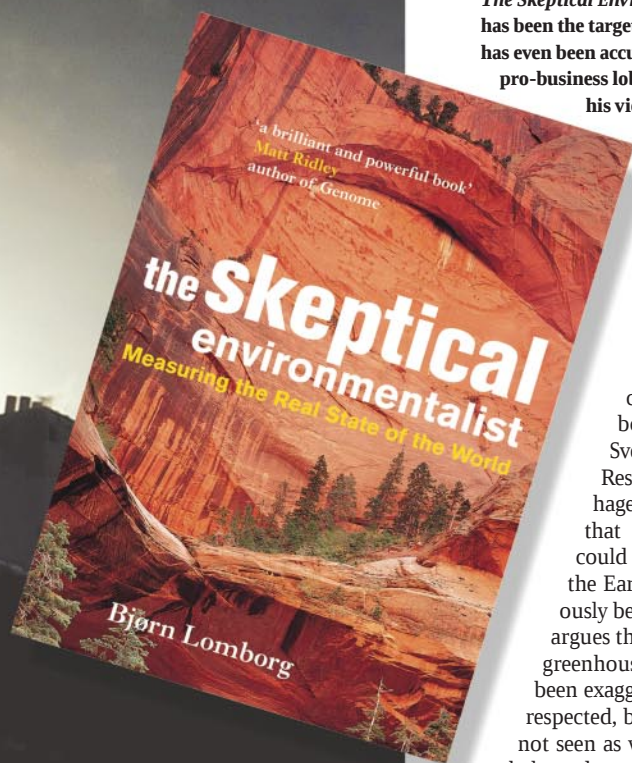
Lomborg formed a study group among his students to take up Simon's challenge. Describing himself as having a left-wing frame of mind, Lomborg says he assumed that it would be easy to debunk Simon's views as merely the product of conservative American thinking. "We all thought it would just be a matter of how much fun we would have showing he was wrong," he says. But Lomborg says the study group ended up agreeing with many of Simon's claims. Excited by what he was finding, Lomborg persuaded *Politiken*, a left-leaning Danish daily newspaper, to

publish four essays summarizing his findings. Those essays evolved into a book that was released in Denmark in 1998 and eventually published in English by Cambridge University Press as *The Skeptical Environmentalist*.

Lomborg's approach was to use a mass of statistics on issues from species extinction to air pollution, taken from authoritative sources such as United Nations agencies, to gauge the state of the global environment. He concluded that things are not as bad as environmentalists have led us to believe. And although he was careful to say that we must continue to work to protect our environment, he also raised questions about whether current initiatives, such as the Kyoto Protocol on climate change, are the most cost-effective way to do so.

This call of 'crisis, what crisis?' resonated with elements of the conservative media: favourable reviews soon appeared in *The Economist*, *The Wall Street Journal* and British newspaper *The Daily Telegraph*. Not surprisingly, the environmental movement reacted with outrage. The World Resources Institute, an environmental research and policy group based in Washington DC, published a list of "Nine things journalists should know about *The Skeptical Environmentalist*" — arguing, among other things, that Lomborg was selective in the studies that he cited, and lacked the credentials to carry out his analysis correctly.

Since the publication in 2001 of his book *The Skeptical Environmentalist*, Bjørn Lomborg has been the target of vitriolic criticism and has even been accused of pandering to the pro-business lobby. But he maintains that his views are based purely on a dispassionate analysis of available data.



tics but in his selection and interpretation of data. "The book is strong on numbers but weak on analysis," Grubb says.

In the chapter on climate change, for example, Lomborg chooses to cite Henrik Svensmark at the Danish Space Research Institute in Copenhagen, whose research suggests that variation in solar activity could have a greater influence on the Earth's climate than has previously been acknowledged. Lomborg argues that the warming influence of greenhouse gases may therefore have been exaggerated. Svensmark is widely respected, but his provocative ideas are not seen as well established. "It wasn't a balanced account," says Grubb. "Lomborg chose to highlight one end of the uncertainty."

Perhaps the most significant criticisms of *The Skeptical Environmentalist* come from those who ought to support Lomborg's statistics-driven approach. Dan Esty, an expert in environmental law at Yale University in New Haven, Connecticut, is a prominent advocate of the idea that carefully designed and measured indicators of environmental health can be used to assess the impact of problems such as water pollution. It sounds as if he should be one of Lomborg's biggest fans. "But the sad truth is that what Lomborg got right was lost among what he got wrong," Esty says.

False economy?

Throughout the book, for example, Lomborg argues that environmental benefits will accrue from increasing prosperity. Esty believes that Lomborg has oversimplified this connection. "I think there is some correlation between economic development and better environmental results," says Esty. "But the suggestion that environmental development stems from economic development is a misunderstanding." Indices of environmental health, Esty points out, show that some developing countries, such as Costa Rica, perform well, whereas other, richer nations, including Belgium, score badly. "There are a great number of policy choices to make and priorities to set at whatever level of economic development a country finds itself in," argues Esty.

Now that the initial furore has died down, most researchers seem to agree broadly with Esty and Grubb: Lomborg has made some

interesting points, but far more thorough analyses of the problems that he tackled can be found elsewhere. Given that other researchers have spent decades addressing the same issues, this conclusion is perhaps unsurprising. And it isn't as if Lomborg is the first author to cast doubt on the claims made by environmental campaigners. So why did the book generate such a violent reaction?

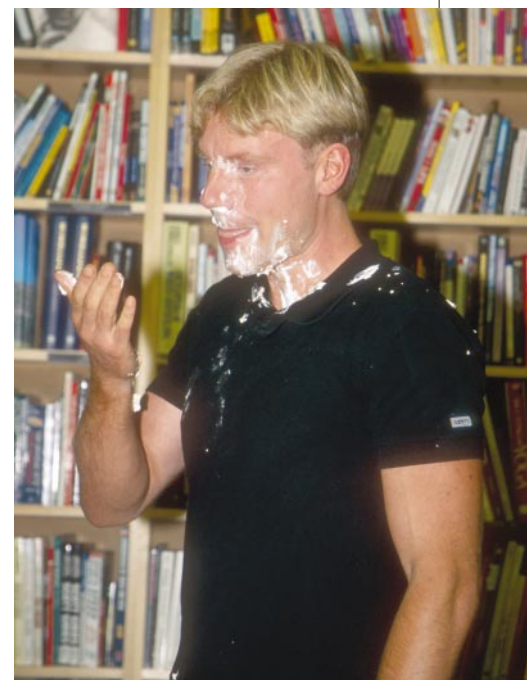
In part, the outcry may reflect an element of panic about the way in which the conservative media seized upon Lomborg's book. Many campaigners feared that his arguments would be used as justification by politicians who oppose environmental legislation — a worry that was heightened in the United States by the Bush administration's apparent disdain for environmental issues. "The book was getting very positive reviews," says Allen Hammond, a senior scientist at the World Resources Institute. "We wanted to warn people that we had problems with it."

But Lomborg's background and character were also important factors. Simon and most other prominent critics of the green movement have been right-wingers, preaching to those who are already ready to reject environmentalists' claims. Lomborg, however, comes from one of Europe's most liberal nations and was even a member of Greenpeace as a student. What's more, he has an engaging manner — exemplified by his good-natured response to the Oxford pie protest — and a rare ability to quote statistics without losing his audience. "He has a verbal and mathematical sharpness — you don't often get that combination," says Toger Seidenfaden, editor-in-chief of *Politiken*.

Campaigners in Oxford, UK, set up an anti-Lomborg website to bring together the burgeoning mass of critical comments. And in September 2001, before a debate at an Oxford bookstore, one activist pushed a pie into his face. "I was stunned," says Lomborg. "But at least the pie tasted good."

Environmental scientists also weighed in against Lomborg. In January 2002, *Scientific American* carried sharply critical articles by four environmental researchers. An accompanying editorial note described the book as a "failure". *Nature* had earlier published a scathing review by Stuart Pimm, a conservation biologist at Columbia University in New York, and Jeff Harvey, an ecologist at the Netherlands Institute of Ecology in Heteren, accusing Lomborg of ignoring research that failed to support his view and of referring to secondary sources, rather than the primary literature (S. Pimm & J. Harvey *Nature* **414**, 149–150; 2001).

Some of Lomborg's critics entered into a series of rebuttals and counter-rebuttals that soon became bogged down in a mire of statistics. "Some discussions have got lost in the details of the details," says Michael Grubb, a specialist on climate change and energy policy at Imperial College, London. But, like many observers, Grubb believes the real weakness of Lomborg's book lies not in sloppy statis-



Pie-eyed: this custard-based assault on Lomborg epitomizes the ire that his views have provoked.

PA, INSET, CUP

H. WARWICK



A. KIM/POL

Trash cans: Lomborg's institute has questioned Denmark's devotion to aluminium recycling.

Some of Lomborg's opponents would dearly love to be able to portray him as a stooge of the political right. But that won't be easy. He seems entirely genuine about his stated position of having formed his views simply through unbiased statistical analysis. "I believe I have looked at the important indicators," says Lomborg. "If I sat down with dispassionate researchers, most of the time we would come up with conclusions that are close to mine."

Lomborg does, however, acknowledge that his position has provided succour for polluting industries, and for right-wing groups that are opposed to environmental legislation. "I know people use me for their political ends," he says. And he has been criticized for his decision in October 2001 to speak to members of the US Congress at a briefing organized by the Cooler Heads Coalition, a Washington DC-based group that campaigns against the Kyoto Protocol. Lomborg argues that his message is the same, whatever the circumstances. "I say the same things at oil-company meetings and ecology meetings. If researchers refrain from saying things that could be used politically, then they start acting as politicians," he says. Nevertheless, he has turned down some offers, including one from a plastics-industry organization that wanted to sponsor a lecture tour of the United States.

So what, in the long term, will be Lomborg's influence on the debate about the state of the planet? In Denmark, at least, he is continuing to make waves through his role as director of the country's Environmental Assessment Institute, which was established

in January 2002 by the country's newly elected centre-right government. With 17 researchers and an annual budget of US\$2 million, its primary aim is to conduct cost-benefit analyses of environmental issues that are important to Denmark.

The institute has, for example, already challenged environmental advocates by questioning whether the money invested in recycling aluminium cans is well spent, and is now working on issues such as soil pollution and waste incineration.

More generally, few would argue that Lomborg's book has won over many hearts and minds. The Bush administration's policies on Kyoto and other environmental issues were largely formed before the book was published, and green activists continue to campaign using the same arguments as before. But some experts argue that the affair contains valuable lessons for environmental scientists: specifically, they argue that it shows how counterproductive it can be to respond to misleading claims with anything other than reasoned scientific argument. The tone of the scientists' attacks on Lomborg was often emotive, and they were sometimes seen as political. "I've yet to see a peer-reviewed response to Lomborg," points out Roger Pielke, an expert on science and technology policy at the University of Colorado at Boulder.

Pimm and Harvey's *Nature* book review, for instance, contains the following statement: "The text employs the strategy of those who, for example, argue that gay men aren't dying of AIDS, that Jews weren't singled out by the Nazis for extermination, and so on." And the series of critiques in *Scientific American* is subtitled: "Science defends itself against *The Skeptical Environmentalist*", as if science itself was under attack.

Gripe hype

Presenting Lomborg as an enemy of science may indeed have been hyperbole, and it certainly seemed to boost his profile. Chris Harrison, publishing director for social science at Cambridge University Press, who handled the book, says that sales quadrupled in the month following the appearance of the *Scientific American* articles — although he stresses that stirring up controversy was not a deliberate marketing strategy.

Perhaps the most surprising development in the affair was the move by a group of scientists, including Harvey, to report Lomborg to the Danish Committees on Scientific Dishonesty, an official body that is responsible for examining accusations of scientific misconduct. Even more surprising, to some observers, was the committees' decision to investigate. In a highly confusing judgement released in January, the committees' deemed *The Skeptical Environmentalist* "to fall within the concept of scientific dishonesty", because of its allegedly biased presentation of data, although the report conceded that there was no evidence

that Lomborg had actually intended to deceive his readers (see *Nature* **421**, 195 & 201; 2003).

The investigation has been heavily criticised for relying on published critiques of the book, in particular the *Scientific American* articles. Lomborg issued a lengthy rebuttal on his website and lodged complaints about the investigation with the Danish parliamentary ombudsman and the government. "He never, ever said anything but what he believed the data showed," says Kenneth Thue Nielsen, a member of Lomborg's original study group who was until recently a researcher at the Environmental Assessment Institute.

Nevertheless, Lomborg loses a little of his calm when discussing the investigation — it is evidently one of the few events over a turbulent couple of years that have really rattled him. "They simply said that if the four critical scientists in *Scientific American* said I was an idiot, then I must be," he complains. Fearing that the publicity surrounding the ruling was damaging the Environmental Assessment Institute's standing, its board of governors launched an independent review of all of the reports it has produced. Board members hope that the review, which should be ready in August, will provide a vote of confidence.

The real loser from the incident may, however, be the Danish Committees on Scientific Dishonesty, which now faces a review of its remit by the Danish government. And the committees' report definitely served to propel Lomborg and his controversial ideas back into the headlines at a time when interest had at last begun to wane. Whereas some news articles simply said that Lomborg had been found guilty of misconduct, others portrayed him as the victim of a witch-hunt.

Harvey has taken part in public debates with Lomborg on several occasions, and his opinions on the book have not changed. But looking back, he acknowledges that the venom with which Lomborg was attacked may have been counterproductive. "The affair has taught me to be more calm and measured," says Harvey. "We should have let the empirical evidence undermine Lomborg."

Harvey's comment illustrates an important lesson to be learned from the affair. *The Skeptical Environmentalist* is packed with facts and figures, yet it was the emotional response that it inspired that will be best remembered. The whole controversy, Pielke laments, is now perceived to have been about politics rather than science, and everyone has been tarred with the same brush: "Scientists are seen as the same as everyone else." ■

Jim Giles is *Nature's* associate news and features editor. Additional reporting by *Nature* intern

Hannah Hoag.

Lomborg's website

♦ www.lomborg.com

Anti-Lomborg

♦ www.mylinkspace.com/lomborg.html

Danish Environmental Assessment Institute

♦ www.imv.dk

Mock turtles

One reptile enthusiast, working with a handful of academics, has described a clutch of new Asian turtle species since the late 1980s. But are they what they seem? Rex Dalton reports on a herpetological débâcle.

They began emerging in the late 1980s: new species of freshwater and land turtles from the waterways and forests of China and Southeast Asia. By the late 1990s, more than a dozen had been described. And that posed a challenge for conservation biologists, who were already struggling to protect the region's existing turtles, which are threatened by poaching for food and traditional medicine. Conservation plans, it seemed, would have to be broadened to include the new discoveries.

Most of the new species hadn't made their scientific debuts in the usual way, by being discovered in their natural habitats by field biologists and described in the context of their surroundings. Ten of them had been bought by a New York veterinarian, William McCord, from two Asian reptile dealers, who said that they had acquired them in markets or isolated villages. From 1987, McCord and his academic collaborators began describing the specimens in specialist journals¹⁻¹⁰ (see Table, overleaf).

Subsequent events have put the status of those specimens in serious doubt. In 1998, McCord's two dealer contacts were named in a US criminal indictment as members of the world's biggest ring of reptile smugglers. Shortly afterwards, an enterprising Californian graduate student began an inquiry that seems to have blown McCord's scientific claims apart. It now seems that six of the species described by McCord and his collaborators are hybrids that were possibly bred in captivity, and three are species that have been described under other names elsewhere in the literature. Only one seems to be unambiguously a new species.

Herpetologists might have been able to dismiss the experience as a mere embarrassment were it not for the conservation implications. Four of the probable hybrid turtles have already been named in the *China Red Data Book of Endangered Animals*, an official national handbook for wildlife conservation, as species that need study and possible protection. "There are very limited resources for turtle conservation in Asia," says James Parham of the University of California, Berkeley, the graduate student who highlighted the problems with the turtles' identities. "It would be extremely unfortunate if precious funds are wasted on hybrid species described from turtle farms."

Turtles and tortoises are struggling for sur-



The taxonomy trail: investigations by James Parham and Haitao Shi (inset) revealed that farms in China are generating hybrid turtles that are very similar to animals that have been described as new species.

vival worldwide. Of the 90 or so species in Asia, the survival of about 75% is threatened. Given the size of the problem, the furore surrounding McCord's papers was the last thing that turtle conservationists needed. "This is very serious," says James Spotila, a herpetologist at Drexel University in Philadelphia, Pennsylvania, and a member of two World Conservation Union panels responsible for assessing the status of turtles. "There are obviously a lot of problems that need to be cleared up."

Private investigations

Until the current débâcle, McCord moved smoothly in herpetological circles. Indeed, he has published extensively on turtles in reputable journals. At his private East Fishkill Animal Hospital in Hopewell Junction, north of New York City, McCord maintains a collection of hundreds of turtles that rivals national facilities.

Some of these specimens were supplied by Keng Liang 'Anson' Wong, a dealer who was based in Penang, Malaysia, and Yuk Wah 'Oscar' Shiu from Hong Kong. Wong is now



languishing in a US jail, serving a six-year sentence. Over the course of nearly a decade, US Fish and Wildlife Service agents coaxed the smuggling ring that Wong headed into sending rare reptile species to the United States. Wong was eventually lured to Mexico, where he was arrested. After fighting extradition for two years, he agreed to face US justice, pleading guilty in 2000 to 40 counts of reptile smuggling and money laundering. Four US accomplices were also convicted, and federal agents are now trying to extradite Shiu from Hong Kong. Records from the Grand Jury indictment in San Francisco suggest that Shiu provided \$45,000-worth of reptiles for

C. FELDMAN

J. PARHAM

shipment to the United States. Shiu did not respond to telephone messages left at his Hong Kong animal shop.

McCord stresses that the turtles he received were imported legally. Given that they were thought to be new species, they would not have fallen under the remit of the Convention on International Trade in Endangered Species of Wild Fauna and Flora, which regulates the wildlife trade. McCord also insists that his endeavours were for science and the preservation of turtles. He now acknowledges that he may on occasions have been “tricked” into buying farm-bred hybrids. But he is holding out hope that some of the new turtles are “natural hybrids”, having crossed in the wild on their own.

McCord still speaks positively of Shiu, with whom he collaborated in 1997 on a documentary film about the threats facing Asian turtles. “He risked his life to expose what is happening,” McCord says. “I don’t believe he intentionally lied.”

Indeed, McCord has continued to publish on specimens provided by Shiu since the allegations of smuggling were made. Last year, McCord was a co-author on a paper¹¹ by researchers at the University of California, Davis, on the genetic analysis of other turtle species. These included specimens collected for McCord by Shiu, reportedly from an isolated Chinese village near the border with Myanmar. Because of McCord’s “long-term professional relationship” with Shiu, the article stated, “we feel confident in the accuracy of the locality data”.

Having delved into other papers in which Shiu provided similar information, Parham doesn’t share that confidence. Parham first began to question the identity of the new species described by McCord and his colleagues while conducting background research for his thesis on turtle palaeontology. “I was just a guy working on fossils,” he says. But as his suspicions about McCord’s

The confused identity of Asian turtle species

Name	First citation	Status
<i>Cuora mccordi</i>	Ref. 2	New species
<i>Cuora chriskearannarum</i>	Ref. 1	Described elsewhere as <i>Cuora pani</i>
<i>Cuora pallidicephala</i>	Ref. 3	Described elsewhere as <i>Cuora zhoui</i>
<i>Cyclemys atripons</i>	Ref. 9	Described elsewhere as <i>Cyclemys pulchriscripta</i>
<i>Mauremys iversoni</i>	Ref. 4	Indicated by genetic tests to be hybrid (<i>Mauremys mutica</i> × <i>Cuora trifasciata</i>)
<i>Mauremys pritchardi</i>	Ref. 10	Indicated by genetic tests to be hybrid (<i>Mauremys mutica</i> × <i>Chinemys reevesii</i>)
<i>Cuora serrata</i>	Refs 7, 15	Indicated by genetic tests to be hybrid (<i>Pyxidea mouhotii</i> × <i>Cuora galbinifrons</i>)
<i>Sacalia pseudocellata</i>	Ref. 6	Probable hybrid (<i>Cuora trifasciata</i> × <i>Sacalia quadriocellata</i>)
<i>Ocadia glyphistoma</i>	Ref. 8	Probable hybrid (<i>Ocadia sinensis</i> × <i>Mauremys annamensis</i>)
<i>Ocadia philippeni</i>	Ref. 5	Probable hybrid (<i>Ocadia sinensis</i> × <i>Cuora trifasciata</i>)

identification of the specimens began to grow, Parham started to raise questions with other herpetologists — and began to hear dark rumours about turtle hybrids.

So began a two-year sojourn into the netherworld of the Asian turtle trade. Largely with his own funds, Parham made repeated trips to China, where he teamed up with Haitao Shi of Hainan University. Together, they visited turtle farms in Hainan province, where hybrids are bred for sale as food, medicine and to turtle collectors, finding specimens that matched the descriptions of the new species identified by McCord. The farms’ proprietors acknowledged selling hybrids to dealers. Finally, in 2001, Parham and his colleagues published two papers^{12,13} detailing their findings, which included genetic analyses strongly suggesting that two of the species described by McCord are hybrids. Separate genetic analysis by a German group revealed that a third of McCord’s specimens also seems to be a hybrid¹⁴.

Identity crisis

“Parham is the one person who pulled all this out in the open,” says Carl Ernst, a herpetologist at George Mason University in Fairfax, Virginia, who collaborated with McCord in the late 1980s, identifying two apparently new turtle species¹². One seems to be bona fide, but the other is now thought to be a previously identified turtle. After growing suspicious about the information supplied by Shiu about the localities in which the turtles had been found, Ernst stopped collaborating with McCord shortly after the papers were published. “The changing localities were driving me nuts,” he says.

But when Parham first started talking about his findings at scientific meetings, his diligence in shedding light on the murky world of the turtle trade, and on the description of hybrids as new species in the scientific literature, wasn’t universally appreciated. Parham says that he met with resistance and experienced professional attacks including a series of critical e-mails. And in 2001, an invitation to participate in a World Conservation Union workshop on endangered turtles at Fort Worth Zoo in Texas was suddenly withdrawn.

Many of McCord’s former collaborators have been left

deeply embarrassed by the still-unfolding story. John Iverson, a herpetologist at Earlham College in Richmond, Indiana, named five new turtle species^{3,5,6,8,9} with McCord between 1989 and 1997; a sixth, described by Iverson and McCord as a new subspecies⁷, was later elevated to the status of a new species by other researchers¹⁵. Four of these turtles, including the elevated subspecies, are now thought to be hybrids; the remaining two examples are previously identified species. “I was trying to do good for turtles,” says Iverson. “To some extent, it backfired.” He stopped collaborating with McCord in 1998, as soon as the indictment of Wong’s smuggling ring came out.

Journal editors are now similarly contemplating the fallout of the affair. Richard Sternberg of the National Center for Biotechnology Information in Bethesda, Maryland, is editor of the *Proceedings of the Biological Society of Washington*, which published six of the papers describing the questionable species. He says that the journal will conduct an inquiry of its publications about dealer-supplied specimens to determine “if any fraudulent or illegally gathered data were unwittingly published”.

Meanwhile, genetic studies to uncover the true identity of McCord’s specimens are continuing. And that, unfortunately, means the expenditure of more time, money and effort that could have been devoted to conserving real, wild turtles. ■

Rex Dalton is Nature’s US West Coast correspondent.

- Ernst, C. H. & McCord, W. P. *Proc. Biol. Soc. Wash.* **100**, 624–628 (1987).
- Ernst, C. H. *Proc. Biol. Soc. Wash.* **101**, 466–470 (1988).
- McCord, W. P. & Iverson, J. B. *Herpetologica* **47**, 407–420 (1991).
- Pritchard, P. C. H. & McCord, W. P. *Herpetologica* **47**, 139–147 (1991).
- McCord, W. P. & Iverson, J. B. *Proc. Biol. Soc. Wash.* **105**, 13–18 (1992).
- Iverson, J. B. & McCord, W. P. *Proc. Biol. Soc. Wash.* **105**, 426–432 (1992).
- Iverson, J. B. & McCord, W. P. *Proc. Biol. Soc. Wash.* **105**, 433–439 (1992).
- McCord, W. P. & Iverson, J. B. *Proc. Biol. Soc. Wash.* **107**, 52–59 (1994).
- Iverson, J. B. & McCord, W. P. *Proc. Biol. Soc. Wash.* **110**, 629–639 (1997).
- McCord, W. P. *Chelonian Conserv. Biol.* **2**, 555–562 (1997).
- Engstrom, T. N., Shaffer, H. B. & McCord, W. P. *Biol. Conserv.* **104**, 173–179 (2002).
- Parham, J. F., Simison, W. B., Kozak, K. H., Feldman, C. R. & Shi, H. *Anim. Conserv.* **4**, 357–367 (2001).
- Parham, J. F. & Shi, H. *Asiatic Herpetol. Res.* **9**, 71–76 (2001).
- Wink, M., Guicking, D. & Fritz, U. *Zool. Abh. Staat. Mus. Tierk. Dresden* **51**, 41–49 (2001).
- Fritz, U. & Obst, F. J. *Zool. Abh. Staat. Mus. Tierk. Dresden* **49**, 261–279 (1997).

The ‘golden coin’ turtle (*Cuora trifasciata*), much sought after for food and medicine in Asia, is often crossed with other species to give hybrids that are then sold as the genuine article.



Japanese system buries the individual researcher

The hierarchy ensures that young scientists' grants are absorbed into the lab's budget.

Sir — The Japanese government is wrestling with reform of its university system to enhance international competitiveness (see *Nature* **412**, 364; 2001 and *Nature* **419**, 875–876; 2002), including trying hard to promote the activities of young researchers by providing better financial support. For example, in fiscal year 2002, the Japan Society for the Promotion of Science (JSPS) boosted the value of special grants to researchers under the age of 37 by 50%, and 8,500 young researchers, primarily university staff, each received \$US10,000 on average. Nevertheless, the real issue is whether such grants are effectively distributed to the intended targets. In fact, a substantial amount of this investment is being diverted to other recipients.

In fiscal year 2002, the JSPS provided a research grant of \$8,000 to each of its 4,800 fellowship recipients, including 2,100 PhD students. The grant is supposed to be distributed directly to individuals to facilitate creativity. Nevertheless, most individual grants are absorbed into supervisors' budgets. Consequently, most young researchers cannot use their own funds as they choose.

It is difficult to find documentation of this practice. Although the JSPS applies strict rules for budget expenditure, and recipients must submit detailed accounting reports to the society, their supervisors tell them how to fill in the reports. This is what has happened to me and to almost all JSPS fellowship recipients with whom I have been in contact.

The core of the problem lies in the *daikozasei* or grand-laboratory system, a hierarchical pyramid in which a professor commands all members, including associate and assistant professors, postdocs and students. The system is found in some European countries (including Germany, from which Japan inherited it), but its feudalistic aspect has been reinforced in Japan. Experimental facilities usually belong to a grand lab rather than to a department, unlike the situation in the United States, where associate and assistant professors control their own labs and perform their own research, relying on departmental core facilities for the use of expensive instruments.

Under the grand-lab system, even young assistant and associate professors suffer from diversion of funds. I have heard from several recipients of JSPS grants that their funds were absorbed into their professors' budgets via abuse of financial reports. During a government meeting on

the reform of competitive research funding held in April 2002, the then minister for science and technology policy, Koji Omi, pointed out that grand-lab professors often systematically acquire grants intended for associate and assistant professors, and that such customs interfere with the independence of young researchers.

Because young researchers in Japan are funded by the larger budget of the grand lab, they can never complain about an offending professor. This is a setback routinely experienced by most young researchers in Japan. The restrictive

academic structure of the grand-lab system wastes a substantial amount of investment in young scientists. The government's plan for promoting young researchers will not work effectively until a solution is found to the structural problems deeply rooted in Japanese academic research, where the individual is buried in the system.

Ippeita Dan

Science Communication Japan and
National Food Research Institute,
Tsukuba Science City 305-8642, Japan
dan@nfri.affrc.go.jp

Early work on elephant gait not to be forgotten

Sir — I was fascinated by the beautiful video analysis of elephants' gaits performed by J. R. Hutchinson *et al.* (*Nature* **422**, 493–494; 2003) but surprised to find no mention of Etienne-Jules Marey, who pioneered an analysis of elephant motion, photographically, in 1887 (E. J. Marey & C. Pagès, *C. R. Acad. Sci.* **105**, 149–156; 1887).

It is true that Eadweard Muybridge had photographed elephants in motion a couple of years earlier, but he did not mark up the salient joints as Marey did, and so could not measure vertical movement at the hip or shoulder. Moreover, his use of a battery of 24 separate cameras could not provide a "chronophotograph", as Marey called it, of successive configurations superimposed on a single plate. (Marey secured this by using an ordinary camera with a slotted-disc shutter rotating behind the open lens.) Such photographs, and the diagrams that could be extracted from

them, were necessary if one was to analyse the biomechanics involved.

Marey's elephants (see illustration) were marked in the same way as those of Hutchinson and colleagues. His motion analyses, which were far more probing and extensive than Muybridge's, are described very fully in Marta Braun's *Picturing Time: The work of Etienne-Jules Marey, 1830–1904* (University of Chicago Press, Chicago, 1992).

Oliver Sacks

New York University School of Medicine, Psychiatry
and Neurology, 2 Horatio Street 3G, New York,
New York 10014, USA

Making a song and dance about emotion

Sir — The idea that dancing rids the body of nervous emotion was an old one even in 1953, as pointed out in '50 years ago' (*Nature* **422**, 673; 2003).

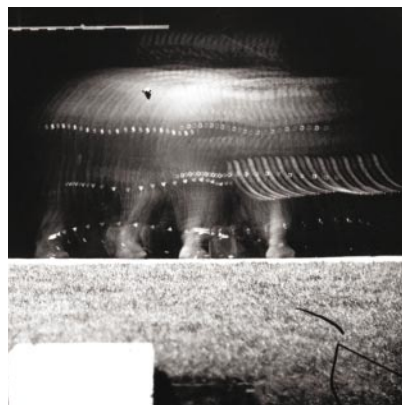
Alfred Wallace, arguing against Darwin's notion of sexual selection in his book *Darwinism* (MacMillan, London & New York; 1889), suggested that "the act of singing [by male birds] is evidently a pleasurable one and probably serves as an outlet for superabundant nervous energy and excitement, just as dancing, singing and field sports do for men".

A man of impeccable manners, Wallace — although not conceding a role for dance in the life of birds — concluded that "singing... may well have originated merely as... an invitation from the male to the female".

I await with nervous anticipation my invitation to *Nature's* next party.

Rupert C. Marshall

School of Biological Sciences, Royal Holloway,
University of London, Egham,
Surrey TW20 0EX, UK



First steps: chronophotographs by Etienne-Jules Marey showed how elephants walked.

The X and Y of sex

The sex chromosomes, especially the degenerate Y, have a lot to answer for.

The X in Sex: How the Chromosome Controls our Lives
by David Bainbridge

Harvard University Press: 2003. 224 pp.
\$22.95, £15.50, E 22.95

Y: The Descent of Men

by Steve Jones

Houghton Mifflin: 2003. 272 pp. \$25
Little, Brown: 2002. 288 pp. £14.99

Jennifer A. Marshall Graves

Just as the mammalian X and Y chromosomes appeared side-by-side some 200 million years ago, so two popular-science books devoted to the two sex chromosomes have appeared almost simultaneously. They are both informative and great fun to read. They tell overlapping stories of how the genetically immaculate X and the sadly degenerate Y chromosomes evolved from an ordinary autosome pair as the Y chromosome degenerated. It's rather like reading the story of a murder from the viewpoints of different participants, including the victim. Both authors detail the consequences to sex, fertility and disease of the single X chromosome's lonely sojourn in males and of the Y chromosome's tenuous hold on existence. And both go further, to explore the cultural spheres of influence of the sex chromosomes.

Contrary to its clear superiority in size and gene content, the X chromosome is dealt with more lightly in David Bainbridge's *The X in Sex*. After a rather roundabout description of the discovery of the X (which I hope will convince students that the X refers to an 'unknown blob' rather than a particular shape) and an approachable (if rather shaggy-dog-like) explanation of sex determination, Bainbridge gets down to explaining why the possession by males of only one copy of the X makes it so special.

In two major chapters, he describes how genes on the X act differently to genes on other chromosomes in causing disease in males (because there is no other X to mask the dud allele) and in inactivating one or other X chromosome in females (to even up dosage between the sexes).

He also summarizes some of the recent work that shows that the gene content of the X chromosome is atypical, and gives an approachable explanation of why you would expect male-advantage genes to hang out on the X. It's too bad that the book was published before the controversial new hypothesis that the X chromosome is smart as well as sexy, containing far more than its fair share of intelligence genes — I'm sure Bainbridge would have had fun with this idea.

In an interesting final chapter, he argues that sex choice is here to stay, and is not necessarily an evil (well, less evil than killing baby girls) — a point of view that I held, too, until I learned of rampages by gangs of mateless and disaffected young men.

I found this book informative and enjoyable, although rather irritating in its many little inaccuracies and convoluted logic (for example, the role of X inactivation is seen as protecting the female against the toxicity of her two X chromosomes, rather than as a secondary effect of the necessity to boost production in the poor male). I also found the book somewhat galling in its reluctance — sometimes even a complete refusal — to present explanations. For instance, Bainbridge is inclined to throw up his hands, saying "no-one can explain" why *SRY* is in such a precarious position on the human Y chromosome, and "I wish I could tell you why" moles have dispensed with the Y altogether. But why didn't he present the interesting, although speculative, explanations for these and several other fascinating observations?

Steve Jones' *Y* is a much more substantial work — an echo, indeed, of Darwin's *tour de force*, after which it is bravely modelled. His point of departure is, again, the degeneration of the Y chromosome, but his theme soon turns to the damage done by the *SRY* gene it bears, which unleashes testosterone on the body and testosterone-fuelled men on the world. There are brilliantly personalized stories of tracking the Y chromosome in human populations, of fertility and mating practices, paternity testing and a thousand other male-directed topics. Jones' easy command of a panoply of amazing factlets puts the book on a par with the wonderful *Ever Since Adam and Eve* by Malcolm Potts and Roger Short (Cambridge University Press, 1999).

The final chapter of *Y* is a provocative discussion of the accumulating problems of being male, from lower sperm counts to rising feminism. Jones rejects Darwin's eugenicist interpretation that patterns of death and mating show the moral superiority



Y so little? Men already carry the smallest chromosome; now their role in life may be shrinking too.

of males and the inexorable weeding out of the unfit, preferring instead the biological testosterone-driven explanation that it is man's sex that makes him weak. One could argue (although I wouldn't) with his dramatic view of men's heritage of death and destruction, and his dark warning that men are expendable (look at Dolly) and could readily fade from view. But his argument is sustained and backed up with serious science, and is worth taking seriously. It reinforces my despairing view that the world would be better off run by females. Or bonobos.

Jones is a wonderful writer. His language is full of colour and wit, his stories lively and to the point. Yet there is depth and clarity in this racy book, and an appreciation, shared with Darwin, Fisher, Haldane and many eminent geneticists today, that sex and sex chromosomes are special and that understanding them requires special rules.

The X and Y chromosomes both have special pulling power (it's called sex) to kindle interest, and they are the chromosomes that perhaps most influence human lives. In the decade in which the fruits of the Human Genome Project regularly put genetics and genomics on the front pages of newspapers, both of these books reach out to audiences far beyond the genetically, biologically or even scientifically literate. ■

Jennifer A. Marshall Graves, Research School of Biological Sciences, Australian National University, Canberra, ACT 2601, Australia.

KOBAL COLLECTION/UNIVERSAL

The Harvard blood group

Edwin J. Cohn and the Development of Protein Chemistry: With a Detailed Account of His Work on the Fractionation of Blood During and After World War II
by Douglas M. Surgenor
Centre for Blood Research/Harvard University Press: 2002. 464 pp. \$34.95, £23.50, E 34.95

Charles Tanford and
Jacqueline Reynolds

In 1920, Harvard Medical School took the unprecedented step of creating a Department of Physical Chemistry, under the leadership of Lawrence J. Henderson, a professor in the parent institution, Harvard University. Members of this department, who had no teaching duties, were charged with the broad mission of conducting research related to medical problems. Henderson appointed as its *de facto* director Edwin J. Cohn, who decided to devote all of its research facilities to the study of protein chemistry.

Henderson had more or less disappeared from the scene by 1927, leaving Cohn in sole charge. Over the next 30 years, Cohn created a laboratory that provided training and sometimes a permanent home for many of the pioneers of protein physical chemistry. In 1950, the department metamorphosed into the grandiosely titled University Laboratory for Physical Chemistry Related to Medicine and Public Health, and Cohn became the first 'university professor', cutting himself free from both Harvard institutions.

After Cohn's premature death three years later, the laboratory underwent several reincarnations, and is known today as the Center for Blood Research, an entity with its own charter and governors, separate from, but affiliated with, Harvard.

Douglas M. Surgenor, who was president of the centre from 1972 until 1987, has written an account of the activities of this research group, with particular emphasis on its work in blood fractionation during the Second World War. He provides extensive discourses on sources of funding and on the machinations of its eccentric director in marshalling not only money, but also scientific collaborators from many other institutions. Every committee that ever met in the quest to provide the US military with an adequate supply of blood products seems to be deemed worthy of mention — the list of names, many of them unknown today, is mind-boggling.

The subtitle of this book is an accurate description of what it contains, but the main title is misleading. This is not a detailed biography of Cohn, although there is some



Increased blood flow: Edwin Cohn set up mobile laboratories to process blood for medical use.

introductory material that gives his family background, together with gratuitously interspersed snippets pertaining to marriages, children and deaths. Interviews with Cohn's surviving family are acknowledged, but events that one might expect to be among the most memorable milestones in the family's history are essentially ignored. For example, the suicide in 1948 of Cohn's wife of over 30 years — pictured throughout the book as consistently helpful to his career — is reported with no explanation of motive or other comment. (Cohn married again within three months of her death.)

Neither is this book a history of protein chemistry, nor even a coherent picture of the place of Cohn's laboratory in the then rapidly developing field of protein studies. Surgenor's interest is with the practical problems encountered in the separation and purification of the proteins of blood serum, specifically the methods and ideas in which he was personally involved, all based on differences in solubility or on centrifugation. But he fails to mention the contemporaneous development in other laboratories of partition chromatography, which would make the Harvard fractionation methods all but obsolete.

Likewise, Surgenor correctly stresses the importance of amino-acid composition in the characterization of individual proteins, and gives credit for this to the laboratory of Erwin Brand at Columbia University (an offshoot of Cohn's overall project). But he doesn't point out that Brand was still using cumbersome microbiological assays for some of the analyses, in which individual analytical

results were based on the growth rate of live bacteria on a Petri dish.

Another illustration of limited perspective is provided by the assertion that Cohn was a leader in the development (or, at least, the encouragement) of X-ray crystallography as a tool for the structural analysis of proteins. But his interest in this field reportedly vanished when he "became involved in a much larger venture": the provision of blood and blood products for military use.

That said, there is much of considerable interest in Surgenor's account of this "much larger venture". In the face of the "burns, burns and more burns" that accompanied the attack on Pearl Harbor, all segments of US society that were involved in providing blood or blood products for therapeutic use laid aside their traditional differences in a single unified campaign. Military officers, meat-processing plants and pharmaceutical companies worked together to convert the academic sector's pilot plant operation at Harvard University into an industrial assembly line. And they apparently put up willingly with the egocentricity and single-mindedness of Cohn and his often abrasive encounters with the minor mortals who inhabited the world around him. Even the somewhat ambiguous nature of the relationship between Cohn and the two great Harvard institutions, the university proper and the medical school, seems to have faded into the background, at least for the duration of the war.

Charles Tanford and Jacqueline Reynolds are at Tarlwood, Back Lane, Easingwold, North Yorkshire YO61 3BG, UK.

Heading for safer shores?

The Biology of Sea Turtles, Volume II

edited by P. L. Lutz, J. A. Musick & J. Wyneken

CRC Press: 2002. 455 pp. \$99.95

N. Mrosovsky

Watching a large turtle emerge from the sea at night to lay its glistening white eggs in the sand is an unforgettable and powerful experience. Sadly, some turtle populations are in serious decline, although others are increasing, leading to debate about whether the risk of extinction has been exaggerated. But biologists and conservationists can all seem to agree on the value of having a better understanding of sea-turtle biology. In the past decade, studies using mitochondrial DNA, satellite tracking and temperature-determined sex ratios have proliferated, as have low-tech beach surveys.

However, too often researchers lack integrative reviews and easy access to carefully archived results. Unrefereed extended abstracts from conferences are often quoted, for want of other sources. The *Marine Turtle Newsletter* has higher standards and sometimes contains important items, but it focuses on alerting readers to developments in their initial stages. So, when the first volume of *The Biology of Sea Turtles* appeared in 1997, it enjoyed considerable success and was frequently cited.

This is now followed by a second volume, which has an emphasis on reproduction (gonadal development, breeding cycles and sex determination). It also covers a variety of other topics, from husbandry to sensory

abilities, and includes thoughtful discussions of growth, age at maturity and population trends. A quarter of the contributions to the volume concern not so much biological as historical, cultural, economic and conservation-oriented aspects of the subject.

Although these reviews are likely to be highly cited, most tend to inform rather than illuminate. For example, a useful chapter on anatomy mentions the highly vascular cartilage of the leatherback turtle. The exciting thing about this work, not captured by a mere reference to the original description, was the creative extrapolation from anatomy to physiology, implying some degree of endothermy in the leatherback. Critical evaluation and guidance about which studies are methodologically best, and why, are often lacking. To fulfil their potential, reviews should be more than the sum of their parts.

Editing seems to have been minimal, resulting in some overlap in text and illustrations. Some photographs are reproduced in both colour and black-and-white, several chapters include similar photographs of gonadal structure, and one photograph, which shows the stages of spermatogenesis, appears twice. It has been said that a picture is worth a thousand words. For pictures illustrating the same thing, 500 each at best would be allowable — make that 250 for photographs lacking scales or resolution.

In contrast, the chapter on migrations and satellite tracking would have been enhanced by including some figures. The editors have not attempted any integration, cross-referencing or explanation of discrepancies. For example, one of the better, more synthetic chapters compares life histories with a mitochondrial-DNA phylogeny for the “seven extant species” of sea turtle, but other chapters assume that east Pacific green

turtles are a separate species, and so end up with eight.

But despite these problems, *The Biology of Sea Turtles* is one of the best places to start — though not to stop — and both volumes may be recommended to libraries and individuals.

N. Mrosovsky is in the Departments of Zoology, Psychology and Physiology, University of Toronto, Toronto, Ontario, M5S 3G5, Canada.

Balancing planets and molecules

Ecological Stoichiometry: The Biology of Elements from Molecules to the Biosphere

by Robert W. Sterner & James J. Elser
Princeton University Press: 2002. 439pp.
\$75, £52 (hbk); \$29.95, £19.95 (pbk)

David W. Schindler

Inspired by the work of Alfred Redfield in the 1950s and 1960s, marine ecologists and geochemists have used stoichiometric principles to guide their studies of nutrient limitation and nutrient cycling in the oceans for almost half a century. Indeed, many of us used ‘Redfield ratios’ for marine algae as the basis for our work on fresh water throughout the 1970s and 1980s. In those days, ‘stoichiometry’ was unknown as an ecological principle, although it was widely used by chemists in calculations involving the masses of reactants or products in chemical reactions.

Little thought was given to developing stoichiometric principles, which were tailored exclusively for other ecosystems until the late 1980s. Robert Sterner and James Elser, along with a small group in Scandinavia, were pioneers in developing these new principles for fresh water. More recently, terrestrial ecologists have jumped onto the stoichiometry bandwagon. Despite more than a decade of intensive scientific work and hundreds of studies demonstrating the importance of stoichiometric principles, there has been no synthesis, until now. So I eagerly accepted *Nature*’s invitation to review this book.

I was surprised by the size of the book. I had expected a small volume attuned to ecological considerations that could be read in a couple of evenings, not 380 pages of text in closely spaced, tiny type. The dense type and occasionally complex language render it a difficult read. Those aged over 50 may have to invest in stronger reading glasses, brighter light bulbs, and eyedrops if they want to read the book in less than a week.

The book provides a clear explanation of the authors’ own work on freshwater systems and the differences in approach of their competitors, and also briefly covers terrestrial



Turning turtle? The risk of extinction may have been exaggerated for some turtle species.

stoichiometry. I hadn't previously thought of applying stoichiometric principles at subcellular or planetary levels, and frankly, after reading about them, I don't think I want to know more. I had also not previously seen the term 'stoichiometry' applied to the coupling of biogeochemical cycles except in connection with limiting nutrients; of course, it fits perfectly.

Few, if any, details of stoichiometry seem to have been overlooked by Sterner and Elser, and their book will be a useful reference tome for many years to come. However, the completeness of the book is also its Achilles' heel. Because of its complexity, I would not give this book to undergraduate ecologists to introduce them to the impor-

tance of stoichiometry. It is a treatise, not a textbook, and represents a massive amount of work.

The hundreds of references in the bibliography are worth the price of the book alone. I would have liked to see a back-referencing system, adding after each reference the page numbers where it is discussed in the text. Such a feature has helped G. E. Hutchinson's *Treatise on Limnology* endure for many decades. Also, the index could be improved; for example, 'trophic cascade' doesn't appear in the index, despite being discussed several times in the book and being one of the most important areas where stoichiometry has enlightened modern ecology.

The proliferation of books costing

US\$100 or more has made it almost impossible for most libraries or scientists to buy more than a small proportion of them. This book is a refreshing exception. It is well edited (although I did find a couple of typographical errors), and figures and tables are clearly reproduced, although there are no photographs. Yet it has a price of only \$29.95. If future authors want their books to be widely read, they should pay attention to these features. All in all, I consider the book a 'must have' for ecologists, limnologists and biogeochemists, and it is an important reference work for others in the Earth sciences. ■

David W. Schindler is in the Department of Biological Sciences, University of Alberta, Edmonton, Alberta, T6G 2E9, Canada.

Science in culture

Behind the wall

Hans Neubert's murals paint a picture of scientific research in East Germany in the 1950s.

Alison Abbott

In 1961, when German artist Hans Neubert was 30 years old, he jumped one of the last trains to west Berlin before the Berlin Wall shot up and sealed so many fates. He left behind in his native Dresden, seemingly forever, a vast number of portraits and landscapes.

He also left behind an extraordinary cycle of ten oil-on-canvas wall paintings at the Friedrich Löffler Institute, now the Federal Research Centre

for Virus Diseases of Animals at Insel Riems, a remote Baltic island north of Greifswald in eastern Germany. Löffler, famous for his 1898 discovery of viruses, founded the institute in 1910, and Neubert's cycle was officially unveiled at its 50th anniversary.

Neubert painted the cycle between 1957 and 1960, when Heinz Röhrer was the director of the institute. Röhrer had initially engaged Neubert to paint his family's portrait, but scientist and artist became fascinated by each other's work. Röhrer let Neubert watch, and sketch, one of his early-morning autopsies of a cow, and the idea emerged of a large painting of the scene for the foyer of the institute's main building. Such was its success that Röhrer commissioned the entire cycle, which now covers the foyer, winds up the stairs and finishes in the gallery on the first floor. Neubert's brief was to depict the full range of the institute's scientific work, from taking blood from rabbits to collecting antibodies, and, as shown here, searching for viral particles by electron microscopy.

A devotee of Rembrandt and Titian, Neubert also absorbed more contemporary influences such as realism and expressionism. Precision was second only to composition in Neubert's priorities. He set up an atelier in the spacious entrance hall of the 1940 building and shrouded it with huge white sheets. Neubert encouraged the scientists to visit him at work to discuss the progress and accuracy of his paintings. Older scientists at the institute still remember his struggle to understand exactly how a hand must be placed to make an injection or to draw blood. The result is a social record at many different levels, as well as an artistic gem of its era.

The only clear deviation from scientific reality occurs in the painting shown here of electron microscopists at work — this 1945 electron microscope would have operated in the dark. But Neubert includes two assistants examining electron micrographs on a light box in the far side of the room. Complementary triangles of different-coloured light

illuminate the faces of each of the five figures and cast long reflections into the floor. The electron microscope, manufactured by Siemens, was one of only two imported into East Germany, and was a source of immense pride to the institute's scientists. At some 4.6 metres high, the painting reflects the instrument's scale and the intensity of the grouped figures.

Artistic realism didn't necessarily equate to political realism, however, and when the cycle was close to completion, Röhrer needed to have them approved by the East German authorities. The paintings were not subversive: they simply celebrated science. But they did not explicitly celebrate the value of that science for the people, as the authorities expected. So Neubert provided a few drafts of a supposed *grande finale*, in which smiling peasants watched scientists vaccinating farm animals. "These pleased the authorities, who were then happy to give approval," recalls Neubert. Approval assured, Neubert destroyed the drafts and painted the final theme that had always interested him: an international group of scientists gathered at the institute to exchange ideas and argue their theories. The 50th anniversary in fact provided exactly such an occasion.

When the Berlin Wall came down in 1989, there were calls for the paintings to be destroyed by those who equated Neubert's style with social realism, a construct associated with the Communist regime. But many fought for their preservation. In 1993, Neubert, by now a contented landscape, still-life and portrait painter in Bavaria, visited Riems again with his wife Barbara, the west Berliner for whom he had risked so much more than 30 years before. He was relieved to find that the colours had held firm against the damp air. Thanks to the protection of the scientists at Riems, Neubert's brief interaction with science has proved equally long-lasting.

Alison Abbott is Nature's senior European correspondent.

A. ABBOTT



Free speech: Hans Neubert encouraged scientists to discuss the accuracy of his paintings.

The person within

Antonio Damasio

If the readers of *Nature* were asked to define the concept of 'self', I imagine that the answers would cluster around two principal meanings. One would be refreshingly precise: "what the immune system identifies as belonging to the body", a topic that was ably tackled by Guy Nossal in an earlier essay in this series. The other meaning, corresponding to the everyday idea of mental self, would be more difficult to pin down. The answers would include: "the sense of one's own being", or "the sum total of qualities that distinguish the mind of one person from that of another", or "one's personal identity".

Can we make the concept of mental self more precise? And are the immunological self and the mental self related to each other? I believe that the answer is yes on both counts.

In an attempt to peer through the conceptual haze of the mental self I have come to consider it from two perspectives: that of introspection and that of biology. Introspection tells me that the mental self is not a thing but a process, one that produces phenomena ranging from the very simple (the automatic sense that I exist separately from other entities) to the very complex (my identity, complete with a variety of biographical details). On the other hand, combining the results of introspection with a biological perspective suggests to me that the mental self represents the individuality and continuity of a living organism.

The value of such a symbol is that it serves as a reference for other contents of the mind, such as representations of the objects with which organisms interact, and of the events in which organisms participate. Having a sense of self as a regular component of the mind allows us to create a freshly minted new entity: a protagonist for the objects and events that populate our mental universe. In effect, the simplest level of self allows us to manufacture the idea that objects and events are perceived from a singular perspective, that of the organism symbolized by the self. At a more complex level, we can generate the idea that the mental processes that occur in this organism are our own property. Finally, with the assistance of past memories of objects and events, we can piece together an autobiography and reconstruct our identity and personhood incessantly.

Where should we look for the neural basis of these self-involving processes? I propose that we search in the neural mappings of our own body, because the body as a whole is the 'thing-process' that is symbolized as the mental self. Others have had this intuition before: Benedict Spinoza and William James perhaps most vividly, but also Friedrich Nietzsche, Martin Heidegger, Maurice Merleau-Ponty and Charles Scott Sherrington, a venerable founder of what we now call neuroscience.

There are two reasons why mappings of the body are well suited to signifying the self in the mind. One is the relative invariance of the body — the design and operations of one's body remain largely the same throughout one's lifetime. The other reason is an often-overlooked fact: that the brain's representation of the structure and operations of the body is continuous.

To illustrate this, look up from the page you are reading and observe intently whatever is in front of you for a few seconds; then look again at this page. As you looked up, many neural stations of your visual system, from the retinas to the cerebral cortex, shifted rapidly from making neural mappings of the page, to mapping the room in front of you. But when you returned to the page these components resumed mapping the page again. In quick succession, the same visual brain regions constructed entirely different neural maps by virtue of the different sensory inputs that you gathered, resulting in different mental images. However, while your visual brain changed obligingly, several regions in your 'body-sensing' brain, which has the job of mapping varied aspects of your body, did not change at all in terms of the kind of object that they represent. The body remained the 'object' of the body-sensing brain all along.

Mental self

Some brain regions can map nothing but the body, and are the body's captive audience. These regions may form the basis of the mind's representation of the 'self'.

The moral of this story is that some parts of the brain are free to roam over the world and to map whatever sound, shape, taste or smell or texture that the organism's design enables them to map. But some other brain parts — those that represent the organism's own structure and internal state — are not free to roam at all; they can map nothing but the body, and are the body's captive audience. It is reasonable to hypothesize that this is the source of the sense of continuous being that anchors the mental self.

We are beginning to discover the neural machinery that is required to map the body. This machinery includes pathways that transmit chemical signals from the internal milieu, through the bloodstream, directly to brain regions such as the area postrema or the hypothalamus; and neural signals from the viscera and muscles that are conveyed by nerve fibres to brain regions in the spinal cord and brainstem. Within the brain itself, dedicated pathways signal this body-related information to certain sectors of the thalamus (a nucleus known as VMpo), and of the cerebral cortex (a sector of the insula). The integration of such signals constructs composite and dynamic maps of the body's state from moment to moment.

These dedicated pathways and regions did not evolve so that we could build a mental self, but rather because continuously updated maps of the body's state are necessary for the brain to regulate life. How intriguing that nature, ever the tinkerer, may conveniently use the same equipment for another purpose: grounding the mental self. Curiously, seen in this light, the mental self becomes a closer relative of the immunological self. ■

Antonio Damasio is in the Department of Neurology, University of Iowa College of Medicine, 200 Hawkins Drive, Iowa City, Iowa 52242, USA.

FURTHER READING

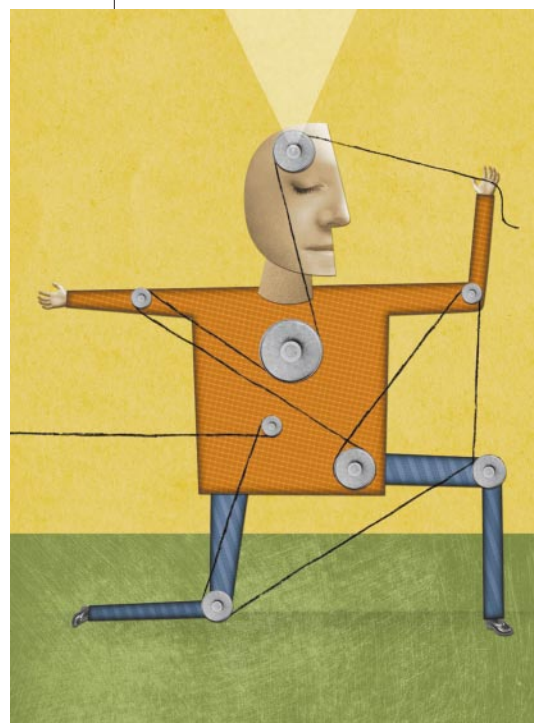
Damasio, A. R. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness* (Harcourt, New York/Heinemann, London, 1999).

Craig, A. D. *Nature Rev. Neurosci.* **3**, 655–666 (2002).

Damasio, A. R. *Looking for Spinoza: Joy, Sorrow and the Feeling Brain* (Harcourt, New York/Heinemann, London, 2003).

Nossal, G. J. V. *Nature* **412**, 685–686 (2001).

J. HESSELBARTH/IMAGES.COM/CORBIS



Sex and self-denial

Ed Newbigin and Richard D. Vierstra

Many plants are self-incompatible — that is, they have mechanisms to prevent fertilization by their own pollen. A familiar and uncommonly versatile protein, ubiquitin, is found to be a central player in one such system.

Ubiquitin is certainly living up to its name. This small protein, first named because of its seemingly ubiquitous presence in animal and plant cells, now seems to be an almost universal regulator molecule. Ubiquitin attaches to a multitude of other proteins, and its well-known role is as a tag that directs unwanted proteins to the machinery that eliminates them. But ubiquitin attachment is also turning out to influence events as diverse as gene transcription and the movement of proteins and other large molecules around a cell^{1,2}. In plants, ubiquitin is known to be involved in regulating responses to hormones and light, in the formation of seeds and flowers, and in the entrainment of circadian rhythms and defence against invading microbes. But there's more. As described in *The Plant Cell*, a group led by Daphne Goring — Stone *et al.*³ — now show that ubiquitin also helps to regulate mate selection in plants.

Much like animals, plants must choose mates from among many suitors. Being hermaphrodites presents them with an additional complication: how to prevent self-fertilization. Left unchecked, self-fertilization reduces genetic diversity and leads to 'inbreeding depression' that, at worst, can contribute to the extinction of a population or species. One mechanism to prevent inbreeding is self-incompatibility. Self-incompatible plants, such as *Brassica napus* (oilseed rape), the species studied by Stone *et al.*, have a high level of variation at a particular genetic location, the S locus. The different versions of the genes at this locus — alleles — enable brassicas to detect their own pollen and thus prevent self-fertilization.

The surface papilla cells of the female reproductive organ, the pistil, actively reject pollen grains (which contain the male gametes) when both have the same S allele, as happens when a plant is pollinated by self pollen. When pollen and papilla cells have different S alleles, the pollen is judged non-self, or compatible, and fertilization can occur. Rejection of self pollen is surprisingly precise: if a compatible pollen grain and a self pollen grain are simultaneously placed on the same papilla cell, only the self pollen is rejected. Rejection means that the papilla cell fails to undergo several localized changes that normally allow a compatible pollen grain to enter the reproductive system and deliver its gametes to the ovary, where the egg

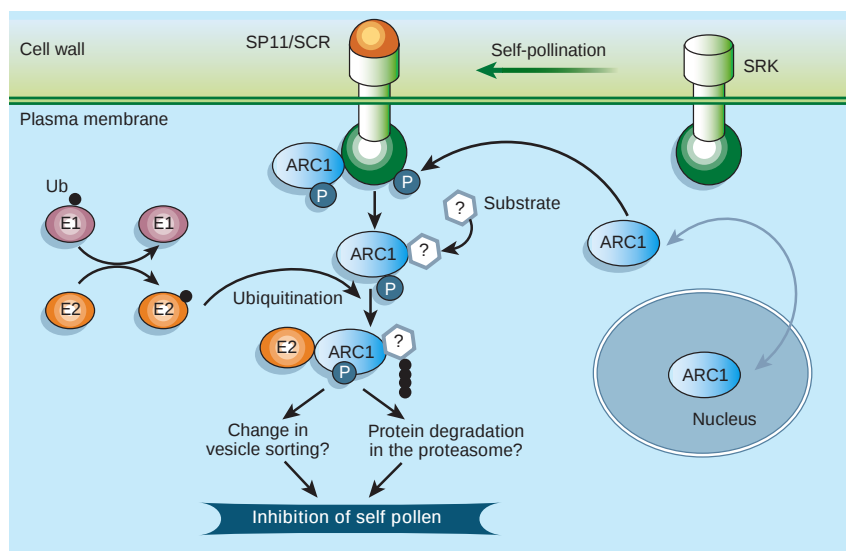


Figure 1 Rejection of self pollen by a brassica papilla cell. In the absence of a self pollen grain, the protein ARC1 shuttles between the nucleus and cytoplasm. In the presence of such pollen, SP11/SCR, released from the pollen coat, binds to and activates SRK, a receptor in the papilla cell's plasma membrane; SRK then phosphorylates (P) both itself and ARC1. Phosphorylated ARC1 in the cytoplasm causes ubiquitin (Ub) to become attached to unknown substrates by means of a conjugation cascade that requires the ubiquitin-conjugating enzyme E2. The experiments of Stone *et al.*³ show that ubiquitination is crucial for self-incompatibility, but how is not yet clear. Ubiquitination might lead to degradation of proteins in the 26S proteasome and their localized depletion, or it might produce a change in vesicle sorting that prevents proteins and other material from being delivered to the pollen contact site. Whatever the case, the upshot is that the papilla cell fails to undergo the changes that allow a compatible pollen grain to enter the plant's reproductive system and eventually perform fertilization.

cells are held⁴. Self-incompatibility in effect slams the door on self pollen.

Self-recognition in brassicas depends on an interaction between two separate and highly polymorphic proteins encoded within the S locus^{5,6}. One product, called SRK, is found on the papilla cell's surface and is a membrane-bound receptor kinase — an enzyme that responds to external signal molecules by phosphorylating, and so activating, internal proteins. The other product, known variously as SP11 or SCR, is a small, soluble protein from the coat of the pollen grain. It is released whenever a pollen grain alights on a papilla cell, but it binds to SRK only when the two proteins are products of the same S allele. Binding of SP11/SCR activates SRK and initiates a kinase-signalling cascade that results in isolation of the incestuous suitor. Up to now, however, the identity of the molecules phosphorylated by SRK and what they do have largely been a mystery.

A few years ago, Goring's group identified⁷ a protein — ARC1 — in brassica papilla cells that could bind to and be phosphorylated by SRK. Plants with lower concentrations of the protein in their pistils had a diminished capacity to reject self pollen⁸. All this pointed to an involvement of ARC1 in self-incompatibility. But how? Scrutiny of the ARC1 sequence revealed a clue. In addition to motifs for localization and protein–protein interactions, there is a region related to the U-box, a domain involved in ubiquitin conjugation. The U-box, like its distant cousin the RING domain, is present in a variety of ubiquitin ligases, or E3s, which are enzymes that perform the final step in the ubiquitination reaction. In these E3s, the U-box/RING domain docks with another enzyme called E2 that carries an activated ubiquitin moiety, while other domains within the E3 select a suitable target. For ARC1, the U-box is near a sequence motif known as an armadillo

repeat that presumably functions in target selection.

In the new work, Stone *et al.*³ show that ARC1 is an E3 and that ubiquitination is crucial for self-incompatibility. They found that ARC1 can direct ubiquitination *in vitro* and that more proteins are ubiquitinated after a self-pollination than after a compatible pollination. An increase in ubiquitin conjugates was not seen in plants with lower concentrations of ARC1. Adding ubiquitin to a protein targets it for destruction by the cell's garbage-disposal unit, the 26S proteasome. Because the application of a proteasome inhibitor to the pistil also impairs the rejection of self pollen, the targets of ARC1 ubiquitination must be degraded by this multi-enzyme complex.

The substrate(s) of ARC1 ubiquitination, and how SRK affects ARC1 activity, remain unknown. Unfortunately, ARC1 is unlike any other known U-box protein, so guilt by association cannot be used to predict its targets. The large increase in ubiquitinated proteins observed after a self-pollination suggests that a marked shift in target selection occurs in the papilla cell. But how does this shift cause the changes seen during pollen rejection?

Figure 1 shows some possibilities. ARC1 might target proteins in the papilla cell that normally promote pollen germination and direct them to the proteasome for breakdown. Degradation must somehow be restricted to near where the papilla cell and self pollen grain touch, so another possibility is related to ubiquitin's role in vesicle sorting: ubiquitination of specific transport proteins helps to direct the flow of vesicles to appropriate cellular destinations¹. Consistent with this notion was the finding by Stone *et al.* that when ARC1 is phosphorylated by SRK it is no longer mostly in the nucleus but enters the cytoplasm, where it seems to associate with the endoplasmic reticulum and secretory system. Focused secretory activity at the contact site and localized loosening of the papilla cell wall are events seen soon after a compatible pollination⁴. ARC1 ubiquitination could prevent the delivery of proteins and other molecules that are essential for germination or wall loosening to the contact site of self pollen.

Stone and colleagues' work constitutes a step forward in our understanding of brassica self-incompatibility. But is there a bigger picture as well? Other plant families have evolved self-incompatibility mechanisms that are radically different to that found in brassicas. Self-incompatible species in the Rosaceae family, which includes apple and some other fruit trees, use a different S-locus-encoded enzyme — a ribonuclease — to inhibit the growth of self pollen. What isn't clear is the S-locus product that identifies self pollen. Two tantalizing reports have shown that a gene encoding another component of

the ubiquitination machinery is near the S-locus ribonuclease gene in the Rosaceae^{9,10}. Possibly, then, ubiquitination helps to control self-incompatibility in both the Rosaceae and Brassicaceae, the family to which brassicas belong. If it does, we shall have to elevate ubiquitin to an even loftier place in the pantheon of plant regulatory molecules. ■

Ed Newbigin is in the Plant Cell Biology Research Centre, School of Botany, University of Melbourne, Victoria 3010, Australia.

e-mail: edwardjn@unimelb.edu.au

Richard D. Vierstra is in the Department of Genetics, 445 Henry Hall,

University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.

e-mail: vierstra@facstaff.wisc.edu

1. Pickart, C. M. *Mol. Cell* **8**, 499–504 (2001).
2. Vierstra, R. D. *Trends Plant Sci.* **8**, 135–142 (2003).
3. Stone, S. L., Anderson, E. M., Mullen, R. T. & Goring, D. R. *Plant Cell* **15**, 885–898 (2003).
4. Dickinson, H. Sex. *Plant Reprod.* **8**, 1–10 (1995).
5. Schopfer, C. R., Nasrallah, M. E. & Nasrallah, J. B. *Science* **286**, 1697–1700 (1999).
6. Takayama, S. *et al.* *Nature* **413**, 534–538 (2001).
7. Gu, T., Mazzurco, M., Sulaman, W., Matias, D. D. & Goring, D. R. *Proc. Natl Acad. Sci. USA* **95**, 382–387 (1998).
8. Stone, S. L., Arnold, M. & Goring, D. M. *Science* **286**, 1729–1731 (1999).
9. Ushijima, K. *et al.* *Plant Cell* **15**, 771–781 (2003).
10. Entani, T. *et al.* *Genes Cells* **8**, 203–213 (2003).

Cosmology

A just-so story

Lawrence M. Krauss

Physicists are learning to live with Einstein's 'fudge factor', the cosmological constant. New thinking attempts to tie its value to other fundamental constants in elementary particle physics.

The recognition, in the light of observational data, that Einstein's infamous cosmological constant might not be zero^{1–3} has changed almost everything about the way we think about the Universe, from reconsidering its origin⁴ to re-evaluating its ultimate future⁵. But perhaps the most significant change in cosmological thinking involves a new willingness to discuss what used to be an idea that was not normally mentioned in polite company: the anthropic principle. This idea suggests that the precise values of various fundamental parameters describing our Universe might be understood only as a consequence of the fact that we exist to measure them. To paraphrase the cosmologist Andrei Linde, "If the Universe were populated everywhere by intelligent fish, they might wonder why it was full of water. Well, if it weren't, they wouldn't be around to observe it!"

The reason that physicists have been so reluctant to consider the anthropic principle seriously is that it goes against the grain. Most physicists have hoped that an ultimate physical explanation of reality would explain why the Universe must look precisely the way it does, rather than why it more often than not would not. Into the fray has entered James Bjorken. In a paper⁶ published in *Physical Review D*, entitled "Cosmology and the Standard Model", Bjorken proposes a new 'scaling' ansatz, based on well-established notions in particle theory, for exploring how anthropically viable a small cosmological constant might be.

The realization that an extremely small, but non-zero, cosmological constant might exist has changed physicists' interest in anthropic explanations of nature precisely because the value it seems to take is otherwise

so inexplicable. In 1996, physicist Steven Weinberg and his colleagues Hugo Martel and Paul Shapiro argued that, if the laws of physics allow different universes to exist with a cosmological constant chosen from an underlying probability distribution, then galaxies, stars and presumably astronomers might not ultimately evolve unless the cosmological constant were not much larger than the one we apparently observe today⁷.

Although this suggestion has spurred several authors to reconsider anthropic arguments, the problem in cosmology is that without a fundamental theory underlying such a probability distribution, very few concrete calculations can be performed. Moreover, while the discussion may centre on fundamental parameters, many of the authors of these discussions are cosmologists, so that little explicit use is made of existing notions from the theory of elementary particles.

However, Bjorken — a particle theorist — notes that a cosmological constant provides a fundamental dimensional parameter that asymptotically characterizes any universe: the so-called de Sitter horizon, R_H (what Bjorken calls R_*). In a universe dominated by a cosmological constant, distant objects recede from an observer at a speed proportional to their distance. Ultimately, beyond a certain, fixed distance, all objects will recede at velocities greater than that of light, and so causal contact will be lost. This distance, the de Sitter horizon, therefore characterizes the effective operational 'size' of an infinite universe undergoing de Sitter-like expansion.

Bjorken suggests that all fundamental dimensional quantities are related to this

de Sitter horizon in a simple way. Taking the Planck scale (where quantum gravitational effects become important, equivalent to an energy of 10^{19} gigaelectronvolts or a length of 10^{-33} cm) as a fundamental scale in nature, Bjorken suggests that other dimensional quantities simply scale in direct proportion to R_H in any universe. In a universe in which R_H is the Planck length, all dimensional quantities, from the grand-unification scale to the mass of the proton, would be identical. As R_H increases, each of these quantities is presumed to have a different scaling factor (in the language of the renormalization group, a different scaling exponent), so that in a universe with R_H equal to that in our currently observable Universe (about 18 billion light years), the different quantities assume the values we measure them to have. With a fixed point at the Planck scale, and a current value, the exponents for each quantity can be inferred.

But it is less clear how very small quantities such as the mass of the electron might scale with R_H . If the small masses are ultimately dependent on small dimensionless coupling constants, they are likely to scale as the logarithm of R_H . But if they are determined by a ratio of some fundamental dimensional scales, they will vary as a power of R_H , according to Bjorken's ansatz. These considerations become important if one wants to use this ansatz to say anything about the likelihood that a universe will assume the values we currently observe it to have. For example, the relative mass of the proton and the neutron and the binding energy of deuterium will depend sensitively on, among other things, the mass of the subatomic particle called the pion.

Bjorken concludes that a surprising number of fundamental constructs in cosmology would be remarkably insensitive to changes in R_H by over 30 orders of magnitude. But he finds that standard nuclear processes associated with stellar evolution — including the famous triple-alpha reaction that allows stars to burn helium to form carbon and is very sensitive to specific beryllium and carbon nuclear configurations — would not be compatible with variations in R_H by more than a factor of about 1.4. This is probably not surprising, because the triple-alpha reaction has long been used in anthropic arguments to constrain variation of the fundamental constants.

In the end, as with so many anthropic arguments, it is hard to know what to make of this result, especially in the absence of any fundamental theory. The suggestion that the small value of the inferred cosmological constant today is tied to the existence of other hierarchies in elementary particle physics through a scaling mechanism is nevertheless quite intriguing as a possible resolution of what is otherwise the most puzzling fine-tuning problem in all of physics. As Bjorken

stresses, perhaps attempts to connect concurrent problems in particle physics and cosmology in this way — even though these types of argument are very speculative — might ultimately provide some guidance for researchers as we try to understand what otherwise seems at present to be a remarkably inexplicable Universe.

Lawrence M. Krauss is in the Department of Physics, Case Western Reserve University, 10900

Euclid Avenue, Cleveland, Ohio 44106-7079, USA.

e-mail: krauss@theory1.phys.cwru.edu

1. Perlmutter, S. *et al.* *Astrophys. J.* **517**, 565–586 (1999).
2. Reiss, A. *et al.* *Astron. J.* **116**, 1009–1038 (1998).
3. Krauss, L. M. & Turner, M. S. *Gen. Rel. Grav.* **27**, 1137–1144 (1995).
4. Steinhardt, P. J. & Turok, N. Preprint astro-ph/0204479 at <http://arXiv.org> (2002).
5. Krauss, L. M. & Starkman, G. *Astrophys. J.* **531**, 22–30 (2000).
6. Bjorken, J. D. *Phys. Rev. D* **67**, 043508 (2003).
7. Martel, H., Shapiro, P. & Weinberg, S. *Astrophys. J.* **492**, 29–40 (1998).

Stem cells

Self-renewal writ in blood

John E. Dick

The ability to self-renew is a defining property of stem cells, and a protein in blood stem cells that controls their self-renewal has been discovered. That same protein is also crucial for the development of leukaemia.

Between birth and death, people produce of the order of 10^{16} blood cells of different types. These specialized cells are continuously produced from precursor cells, which in turn must be replaced by cells further up the blood hierarchy. Ultimately, the entire blood system is fed by a pool of rare haematopoietic stem cells (HSCs)¹. But how does this small pool sustain blood production over a lifetime without being depleted? The answer involves a process termed self-renewal: when HSCs divide, one or both 'daughter' cells can retain the properties of the parent, rather than — as is usual after division — becoming more specialized. Despite some interesting findings, we still know little about how the balance is tipped towards self-renewal². This is not just a biological conundrum: it is also important in developing treatments that involve culturing HSCs *in vitro*. On pages 255 and 302 of this issue, Lessard and Sauvageau³ and Park *et al.*⁴ provide new insight into the problem — they describe a gene regulator that governs the

self-renewal of both HSCs and leukaemic 'stem cells'.

Cell division is an inherently error-prone process, with mutations arising frequently during the necessary replication of DNA. And the more divisions that occur, the more mutations can accumulate, making the development of cancer more likely. Achieving a lifetime of blood production while avoiding a high rate of malignancy is possible only because the blood system is ingeniously organized as a hierarchy. Most proliferation occurs within pools of progressively differentiating precursors that are committed to particular blood-cell lineages. So the likelihood of acquiring enough cancer-predisposing mutations before the final differentiation into non-proliferating, mature blood cells is low.

At the top of the blood-cell hierarchy are the HSCs, from which lineage-committed precursors are generated. When an HSC divides, in theory both daughter cells could become committed to a specific lineage (Fig. 1). But that would result in the eventual

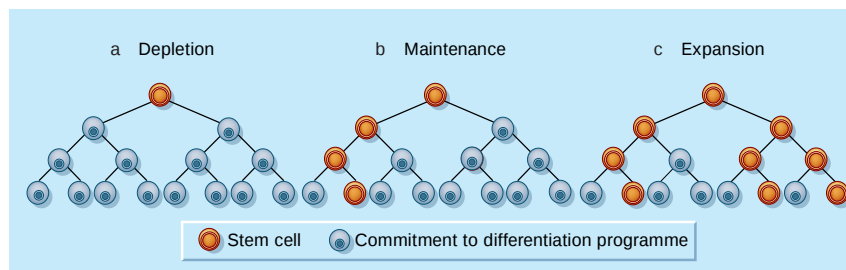


Figure 1 The choices available to haematopoietic stem cells (HSCs) upon division. Daughter cells can either commit to a programme of differentiation that will eventually result in mature, non-proliferating cells, or retain HSC properties. The choice is probabilistic and governed by intrinsic molecular programmes, including, as now shown^{3,4}, the transcription factor Bmi-1. **a**, When both daughters are committed, the clone will become depleted of HSCs and will eventually die out. **b**, During steady-state haematopoietic development, the probability that an HSC will generate at least one daughter cell with HSC properties is about 0.5, and the average number of HSCs will be maintained. **c**, Under some circumstances, especially after transplantation¹⁸, both daughters retain HSC properties, and the overall numbers of HSCs increase. A fourth outcome is that both daughters die.



100 YEARS AGO

In the New Year's number of *Nature* there appeared an account of a basil, *Ocimum viride*, a plant which is known to the natives of Nigeria as a protection against mosquitoes. Captain Larymore, by whom this information was first obtained, in a recent letter to the *Times* mentions that he has brought home a plant which he has presented to the authorities of the Kew Gardens, and that it may be seen there. He also states that the natives believe in its efficacy when taken as an infusion in cases of malarial fever. Further evidence is offered in another letter to the *Times* by Sir George Birdwood as to the knowledge widely spread among the Hindus of these qualities of the basil, which occur wild, and are generally cultivated in India. Thus, during the formation of the Victoria Gardens in Bombay, the workmen were attacked both by mosquitoes and malaria, when upon the recommendation of the Hindu manager basil plants were placed round the gardens, with the result that the unhealthy nature of the locality was effectually changed.

From *Nature* 14 May 1903.

50 YEARS AGO

At La Rocque, in Jersey, the spring tides go out about two miles leaving many large lagoons with sandy bottoms and depths of from two to four feet of water. Here we often spear flat-fish, plaice, soles, etc., sometimes getting as many as twenty in a tide; here also may be found spider crabs (*Maia squinado*) which are often taken in Jersey for food, but which are considered out of season in the autumn. In September last, the catches of flat-fish were down to about half a dozen, owing to the number of octopuses that continually disturbed the bottom and kept the fish moving. The spider crabs had collected into large heaps, about two feet high and three feet in diameter, with their legs so entangled as to make it difficult to separate a crab from a heap. The octopuses captured some from the outside of the masses, but the greater number survived, and day after day the heaps remained. None of the fishermen to whom I have spoken of this behaviour on the part of spider crabs had ever noticed it previously. The octopuses have now (October) moved to deeper water, and the crabs too have disappeared to their winter quarters.

From *Nature* 16 May 1953.

depletion of the HSC pool, so HSCs generally generate at least one daughter that retains stem-cell properties. This ability to self-renew distinguishes HSCs from all other cell types in the haematopoietic hierarchy, and is likely to be a defining property of stem cells in other organ systems as well.

The phenomenon of self-renewal has attained almost mythical status since the 1960s, when Till and McCulloch first identified and quantified HSCs⁵⁻⁷. This early work showed, among other findings, that self-renewal is an intrinsic property of these stem cells. Yet the molecules within HSCs that control self-renewal have remained elusive. Recently, several proteins that are active during embryonic development have been shown to induce self-renewal in adults^{1,2}; these include proteins from the Wnt^{8,9}, bone morphogenetic protein, Sonic hedgehog, and Notch families. But these molecules act on, rather than in, HSCs, begging the question of what lies downstream of them inside the stem cells. One such downstream molecule could be Hoxb4, a gene-transcription factor that, upon overexpression in HSCs, induces significant cell multiplication¹⁰. However, mice lacking this protein do not show defects in HSC self-renewal.

Lessard and Sauvageau³ and Park *et al.*⁴ have now found that Bmi-1, a transcriptional repressor¹¹, is essential to HSC self-renewal. Park *et al.* used gene-expression profiling to identify the *Bmi-1* gene as being highly expressed in HSCs in mice and humans. They, and Lessard and Sauvageau, then transplanted HSCs from the fetal liver or adult bone marrow of *Bmi-1*-deficient mice into normal mice. The *Bmi-1*-deficient HSCs generated a normal pattern of blood cells — but the contribution was only temporary, because of an inability to sustain the precursor pool. Moreover, six weeks after bone-marrow HSCs had been taken from these primary recipients and transplanted into secondary recipients, there were no *Bmi-1*-deficient cells in the secondary mice. So the cells had failed a classic test of self-renewal.

These results, together with the fact that *Bmi-1*-deficient mice die of bone marrow failure within two months of birth¹², indicate that *Bmi-1* is essential for the self-renewal of HSCs. After further gene-expression analysis and some functional studies, Park *et al.* suggest that the effects of *Bmi-1* are mediated through its repression of the genes encoding p16^{ink4a} and p19^{Arf}, proteins that respectively inhibit cell proliferation and enhance cell death.

Lessard and Sauvageau³ also linked *Bmi-1* to the self-renewal of leukaemic stem cells (L-HSCs). There is strong support for the idea that cancer is a stem-cell disease¹³. The best evidence for this comes from studies of acute myeloid leukaemia (AML) in humans, in which most leukaemic cells have limited

proliferative capacity and must be constantly replenished by rare, self-renewing L-HSCs¹⁴. So, like the normal blood system, a clone of leukaemia cells seems to be organized as a hierarchy that originates from a stem-cell pool and retains remnants of the normal developmental programme. It has also been proposed that the initial, cancer-causing ('transforming') mutations in AML occur in HSCs, rather than in lineage-committed precursors (reviewed in refs 15, 16). Self-renewal is a key characteristic of L-HSCs, and so fewer mutations would be required to generate fully leukaemic cells if they were to originate from HSCs (which can already self-renew) as opposed to precursors. This idea predicts similarities in the molecular programmes of normal and AML stem cells.

Lessard and Sauvageau tested this hypothesis by expressing the *Hoxa9* and *Meis1a* genes in mouse fetal liver cells. They had previously shown that, without the need for further genetic changes, these genes reproducibly generate mouse myeloid leukaemia with the same hierarchical organization as the human disease, and that the target for leukaemic 'transformation' in this case does indeed lie in the HSC pool¹⁷. They have now found³ that these genes induced leukaemia in *Bmi-1*-deficient mice with the same kinetics and characteristics as in normal mice, implying that *Bmi-1* is not necessary for the leukaemia to begin and progress. But the L-HSCs from *Bmi-1*-deficient mice could not self-renew, resulting in far fewer leukaemic cells in the blood of primary mouse recipients and an inability to produce leukaemic cells in secondary recipients. This result clearly identifies self-renewal as an essential component of the development of leukaemia, distinct from the potent differentiation block and proliferation induced by *Hoxa9* and *Meis1a*.

The authors³ went on to show that two known targets of *Bmi-1*, the genes encoding p16^{ink4a} and p19^{Arf}, could occasionally become altered in some clones of cultured *Hoxa9*–*Meis1a*-induced, *Bmi-1*-deficient cells. This could to some extent compensate for the loss of *Bmi-1*, resulting in sustained cell production. But the inability of these highly proliferative subclones to initiate leukaemic growth *in vivo* hints that other targets of *Bmi-1* must also be required for self-renewal.

Together, these exciting findings^{3,4} should help researchers to develop methods by which to multiply human HSCs in culture. Moreover, the identification of a gene that is involved in self-renewal in both HSCs and L-HSCs lends strong support to the idea of cancer stem cells, and to the hypothesis that HSCs are the initial target for transformation in some leukaemias. I predict, however, that even lineage-committed cells might occasionally become targets for transformation, if certain cancer-causing mutations restore the self-renewal machinery. Finally,

anti-leukaemia treatments have historically sought to eliminate proliferating cells. But, like normal HSCs, L-HSCs rarely proliferate and so are resistant to such therapies. The identification of L-HSC-specific genetic targets, especially those involved in the enhanced self-renewal seen in L-HSCs, could lead to new treatments that stop the disease at its source. ■

John E. Dick is in the Division of Cell and Molecular Biology, University Health Network, and the Department of Molecular Genetics and Microbiology, University of Toronto, 620 University Avenue, Toronto, Ontario M5G 2C1, Canada. e-mail: jdick@uhnres.utoronto.ca

1. Kondo, M. *et al.* *Annu. Rev. Immunol.* **21**, 759–806 (2003).
2. Devine, S. M., Lazarus, H. M. & Emerson, S. G. *Bone Marrow*

- Transplant.* **31**, 241–252 (2003).
3. Lessard, J. & Sauvageau, G. *Nature* **423**, 255–260 (2003).
 4. Park, I. *et al.* *Nature* **423**, 302–305 (2003).
 5. Till, J. E. & McCulloch, E. A. *Radiat. Res.* **14**, 213–222 (1961).
 6. Siminovitch, L., McCulloch, E. A. & Till, J. E. *J. Cell. Comp. Physiol.* **62**, 327–336 (1963).
 7. Till, J. E., McCulloch, E. A. & Siminovitch, L. *Proc. Natl Acad. Sci. USA* **51**, 29–36 (1964).
 8. Reya, T. *et al.* *Nature* advance online publication, 27 April 2003 (doi:10.1038/nature01593).
 9. Willert, K. *et al.* *Nature* advance online publication, 27 April 2003 (doi:10.1038/nature01611).
 10. Sauvageau, G. *et al.* *Genes Dev.* **9**, 1753–1765 (1995).
 11. van Lohuizen, M. *et al.* *Cell* **65**, 737–752 (1991).
 12. van der Lugt, N. *et al.* *Genes Dev.* **8**, 757–769 (1994).
 13. Dick, J. E. *Proc. Natl Acad. Sci. USA* **100**, 3547–3549 (2003).
 14. Lapidot, T. *et al.* *Nature* **367**, 645–648 (1994).
 15. Wang, J. C. *et al.* in *Hematopoiesis: A Developmental Approach* (ed. Zon, L. I.) 99–118 (Oxford Univ. Press, New York, 2001).
 16. Bonnet, D. & Dick, J. E. *Nature Med.* **3**, 730–737 (1997).
 17. Kroon, E. *et al.* *EMBO J.* **17**, 3714–3725 (1998).
 18. Iscove, N. N. & Nawa, K. *Curr. Biol.* **7**, 805–808 (1997).

Genomics

Yeast rises again

Steven L. Salzberg

What's in a genome? The short answer is that you can't really say in detail for any one species until you have the genome sequences of a variety of other species — some closely related, others less so — to compare it with.

While the human and mouse genomes have lately dominated public discussions of genome science, yeast researchers have quietly continued to forge ahead with analysing *Saccharomyces cerevisiae* and related yeast species. Yeast has long been a favourite of biologists, who use it as a model for investigating the biology of higher organisms. Although yeasts are unicellular, like bacteria, they have a cell nucleus, putting them in the biological kingdom of eukaryotes (Eukaryota), a group that includes humans. Yeast was the first eukaryote to have its genome completely sequenced, in 1996¹. And yeast has long been a part of human commerce and culture, being used in brewing and baking since ancient times². *Saccharomyces cerevisiae* itself was first isolated in beer in 1837.

On page 241 of this issue, Kellis and colleagues³ describe the results of sequencing three more yeast species — *Saccharomyces paradoxus*, *S. bayanus* and *S. mikatae* — and comparing them with *S. cerevisiae*. Two of these species, *S. bayanus* and *S. paradoxus*, are used in winemaking, but the three new species were chosen mainly because they are evolutionarily very close to *S. cerevisiae*, and the implications of the study go well beyond brewing and viniculture. The analyses will produce a substantial revision in our knowledge of the yeast genome, and provide strategic directions for how we might select other sequencing targets to advance understanding of the human genome.

The original *S. cerevisiae* project was a model of international collaboration among small laboratories: 55% of the genome was

sequenced by a network of European labs, with the rest of it being generated at five large (at the time) centres. More than 600 scientists participated in the project, sequencing a cosmid (a 40-kilobase fragment) at a time and finishing the sequences to a high level of accuracy with minimal automation¹. In the seven years since then, genome sequencing has moved increasingly into large-scale factory operations: the three new genomes were produced at a single centre using the whole-genome shotgun strategy, which is much more rapid than the earlier piecemeal sequencing approaches.

The capacity of today's largest centres allows them to sequence a yeast genome — 12 million base pairs or so — in just a few days of full-scale production. Developments

in genome assemblers^{4,5} make it possible to assemble the fragments produced by the whole-genome shotgun approach into relatively large pieces with few mistakes. This assembly step works even better when a reference genome already exists: the assembled fragments can be quickly mapped onto the reference, assuming that relatively few rearrangements have occurred since the species diverged. As shown in Fig. 2 of Kellis and colleagues' paper (page 244), all 16 chromosomes from each of the three newly sequenced genomes map beautifully onto *S. cerevisiae*.

The main reason for sequencing the three further yeast species is to better understand *S. cerevisiae*. A few years ago, Cliften and colleagues⁶ and the French consortium Génolevures² independently conducted studies that demonstrated the power of sequencing close relatives of *S. cerevisiae*. Both groups recommended full-scale sequencing of additional yeast species, which they argued would result in the discovery of noncoding RNA genes, genes encoding small proteins, and gene-regulatory sites. In a spot-on prediction, Cliften *et al.* estimated that at least 40 small proteins (shorter than 100 amino acids) would be identified by comparing *S. cerevisiae* with *S. bayanus*, *S. mikatae* and a third species. Kellis and colleagues found 43.

Comparative genome sequencing also helps to resolve the surprisingly difficult question of just how many genes there are. Genes — those parts of a DNA sequence that encode proteins or functional RNA molecules — are the basic components that do all the work of a cell, from reproduction to metabolism to inter-cell communication. Nearly seven years after the yeast genome was first published, the gene count stands at approximately 6,128 (refs 7–9), only 150 or so fewer than the initial estimate. But many of these genes are still questionable: the Génolevures programme¹⁰, after analysing sequences sampled from 13 different yeast

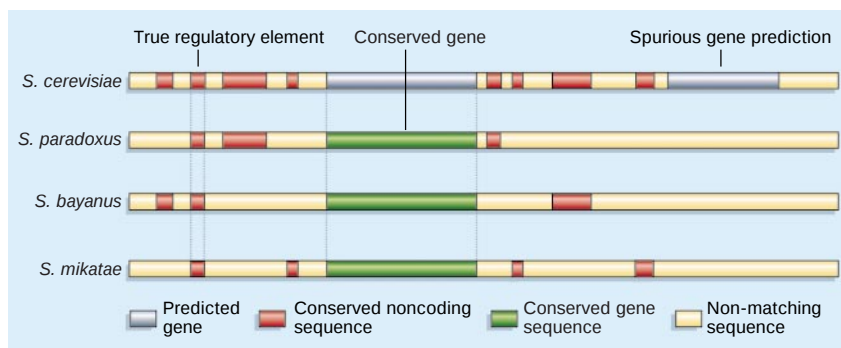


Figure 1 **Comparative genomics.** Comparing the DNA sequences from several species makes it possible to find regulatory regions — short sequences that turn genes on and off — and eliminate spurious gene predictions. Red boxes highlight areas of sequence similarity between at least two species. Functional sequences — genes and regulatory elements — tend to be conserved across all species. The figure shows how one true regulatory element and one correctly identified gene might emerge from a comparison of four yeast species.

species (including *S. bayanus*), and failing to find matches for 742 genes, estimated that the gene count should be reduced to 5,651. The crux of their argument was the assumption that if a gene is functional, it should be conserved among closely related species (see Fig. 1). Because the Génolevures project covered only 20–40% of each genome, the possibility remained that those ‘missing’ genes might be found in the unsequenced regions.

Kellis and colleagues³ have now closed the door on that possibility: their sequence data cover 98% of two species and 93% of the third. Based on sequence alignments among the species, they largely confirm the results of the earlier study and argue that 503 genes should be deleted from the yeast catalogue, leaving 5,726 genes, of which 43 are newly discovered in their study.

Regulatory sequences, which sit outside genes and turn them on and off, are the key to understanding how a genome fits together. Whether we are comparing human and mouse, or yeast and yeast, we still have to answer the puzzling question of how seemingly huge differences in physical, biochemical or behavioural characteristics can result from sometimes tiny differences in the protein sequences. Regulatory sites occur virtually anywhere in the vast areas between protein-coding regions, and they can be identified because — unlike non-functional regions — they are conserved. The closer two species are, the more regulatory sites they are likely to share. Unfortunately, if the species are close enough, many pieces of non-functional DNA will be conserved merely by chance. A nice solution to this problem is to sequence more than two related species, dramatically increasing the signal-to-noise ratio (Fig. 1). The idea is that functional sequences should be conserved across multiple species, whereas chance conservation will only appear in pairwise comparisons. Using this principle, Kellis and colleagues have identified 42 novel sequence motifs that appear likely to have biological functions in yeast.

The most dynamic parts of the yeast genome are the chromosome ends, called telomeres. Kellis *et al.* aptly describe the rapid change and exchange going on in these regions as “genomic churning”. The telomeres contain many genes not found elsewhere in the genome, and it appears that they form a crucible in which genomic change occurs: as the telomeres swap back and forth between chromosomes, they can carry pieces of genes along with them, which may combine with others to create new genes. Similarly rapid changes are evident in the telomeres of the malaria parasite *Plasmodium falciparum*¹¹. Given the important events associated with these regions, sequencing of telomeres of the human genome — which, so far, have been neglected — should become a priority.

If 8% of the estimated protein complement

of yeast was wrong, how much of the human counterpart might be eliminated by a similar study of mammalian species? How many genes and regulatory regions would be discovered? The human gene count has already dropped from more than 31,000 to just under 25,000 (refs 12, 13) in the two years since the initial genome publication. To understand our own genome better, we should sequence the genomes of several other mammals besides mouse and rat (both are almost done), choosing species that vary in their evolutionary distance from humans. At the chromosome level, only primates have as much structural similarity to human as these four yeast species share, and even those have different chromosome numbers: chimpanzee (*Pan troglodytes*) has 24, gibbon (*Hylobates concolor*) has 25, and macaque (*Macaca fuscata*) has 21. At the sequence level, we might gain more information from studying more distant mammals, such as cat (*Felis catus*), pig (*Sus scrofa*) and dolphin (*Tursiops truncatus*).

This new study of yeast genomes³ makes

it clear that comparative genome sequencing has tremendous analytical power: it offers the prospect of enhancing our knowledge of thousands of genes at once as well as providing fresh clues about the function of the vast amount of genomic DNA that does not encode genes. As it has done before, lowly yeast shows us a path towards a better understanding of our own biology.

Steven L. Salzberg is at The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA.

e-mail: salzberg@tigr.org

1. Goffeau, A. *et al.* *Science* **274**, 546, 563–567 (1996).
2. Feldmann, H. *FEBS Lett.* **487**, 1–2 (2000).
3. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. *Nature* **423**, 241–254 (2003).
4. Myers, E. W. *et al.* *Science* **287**, 2196–2204 (2000).
5. Batzoglou, S. *et al.* *Genome Res.* **12**, 177–189 (2002).
6. Cliften, P. F. *et al.* *Genome Res.* **11**, 1175–1186 (2001).
7. Snyder, M. & Gerstein, M. *Science* **300**, 258–260 (2003).
8. <http://genome-www.stanford.edu/Saccharomyces>
9. <http://mips.gsf.de/proj/yeast/CYGD/db>
10. Malpertuy, A. *et al.* *FEBS Lett.* **487**, 113–121 (2000).
11. Gardner, M. J. *et al.* *Nature* **419**, 498–511 (2002).
12. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
13. www.ensembl.org

Molecular biology

Disruptive influence

Marco Foiani

Recombination is a vital cellular process implicated in DNA metabolism — but it must be tightly controlled. The discovery of a protein that disrupts recombination intermediates sheds light on the control mechanisms.

On pages 305 and 309 of this issue, Krejci and co-workers¹ and Veaute and colleagues² describe a biochemical mechanism that controls the genome ‘shuffling’ occurring in dividing cells and in DNA repair. Their findings have implications for how genome stability is maintained, and hence for the development of cancer.

Genome shuffling is referred to as ‘recombination’, and is a cellular process by which extensive tracts of DNA are moved from one part of the genome to another. There are several recombination pathways³, some not yet well characterized, which are routinely used by normal cells to repair damaged chromosomes, to assist in DNA synthesis, and even to regulate gene expression. Recombination also occurs during the production of eggs and sperm, in which its function is to mix the genetic information such that each egg and each sperm is genetically different.

Despite its importance, however, recombination can sometimes be harmful: it can generate damaging genomic rearrangements, as well as intermediate structures that cannot be processed normally. Cells need to coordinate recombination with other responses to DNA damage, with progression through the cell-division cycle, and with

chromosome replication. Otherwise, cells invariably become genetically unstable as the proteins that bring about recombination take over the chromosomes. During DNA replication, for instance, the double helix unwinds and separates, and the two strands are used as templates to make another helix. Replication frequently stalls, and a ‘check-point’ ensures that the separated DNA (the replication fork) maintains its integrity during these pauses. Without this check-point, abnormal replication intermediates form and are processed by unscheduled recombination⁴. In addition, in some inherited human diseases — Werner, Bloom and Rothmund–Thomson syndromes — mutations in enzymes implicated in DNA metabolism (DNA helicases) cause increased recombination, genome instability and a predisposition to cancer⁵.

So cells must have mechanisms to control recombination and to prevent harmful chromosome rearrangements. One possible mechanism in yeast involves the Srs2 protein (a human relative of which has not yet been discovered, but it is surely only a matter of time). Srs2 is another DNA helicase — it can unwind double helices — and it has previously been implicated in DNA replication, in restarting the cell cycle after DNA-damage-

induced arrest, and in recombination^{6–9}. For instance, mutations in the gene encoding Srs2 lead to excessive recombination⁷. Krejci *et al.*¹ and Veaute *et al.*² now provide biochemical evidence that Srs2 actively inhibits a key step in one particular recombination process — homologous recombination, or genetic exchange between two matching DNA regions.

During homologous recombination, single-stranded DNA must be produced, and the Rad51 protein binds this DNA to form so-called Rad51 nucleofilaments. Rad51 then mediates the exchange of this strand with a complementary tract of DNA. Krejci *et al.* and Veaute *et al.* show that Srs2, as well as acting as a helicase, also has a 'translocase' activity: it dislodges Rad51 from these filaments, thereby preventing recombination.

These findings explain why alterations in Srs2 are associated with hyper-recombination in yeast⁷ (which is reminiscent of the excessive recombination seen in cancer cells). The results might also provide an explanation for other previous findings — and they raise new questions.

For instance, single-stranded DNA might signal the presence of DNA damage^{4,9}, leading to the recruitment of specialized proteins that activate the checkpoint response. The checkpoint then delays the cell cycle, allowing time for the damage to be repaired by various processes, some of which are mediated by Rad51. Srs2 is known to be involved here: it is phosphorylated in response to DNA damage⁸ and, in its absence, cells manifest obvious checkpoint alterations^{8,9}, such as a hyperactive checkpoint that stops the cell cycle from restarting even when the damage has been repaired⁹.

Perhaps the newly discovered inability of Srs2 mutants to dislodge Rad51 from nucleofilaments can explain this aberrant checkpoint: it might be necessary to remove Rad51 after recombination-mediated DNA repair so that the proteins that activate the damage-induced checkpoint can also be removed⁹. Veaute *et al.* also suggest that the Rad51 nucleofilaments themselves could be a checkpoint-activating signal, and hence that the removal of Rad51 by Srs2 is necessary to tell the cell that division can begin again. This is plausible, although cells depleted of Rad51 can still promote checkpoint activation. But regardless of whether they signal to the checkpoint, Rad51 nucleofilaments can clearly form during chromosome repair, making it essential that they be dismantled by Srs2 during recovery.

Srs2 also bears a relationship with Sgs1, the yeast counterpart of the human helicases that are defective in Werner, Bloom and Rothmund–Thomson syndromes¹⁰. When both Sgs1 and Srs2 are mutated, Rad51-mediated recombination causes cell death¹⁰. Although the functional interaction between these two helicases remains unknown, it is possible that

they have a similar role in processing Rad51 filaments. If these proteins do have the same biochemical function, it is not surprising that they can sometimes substitute for each other¹¹. But they do not seem to be redundant, hinting that they might act at different cell-cycle stages or different steps in replication. Perhaps Sgs1 is involved during the replication of damaged chromosomes by antagonizing the formation of Rad51 intermediates and hence promoting specialized, replication-specific repair processes. Srs2, by contrast, could preferentially contribute to the processing of Rad51 filaments after the passage of the replication fork, and later on in the cell cycle.

Another question is whether Srs2's translocase activity is implicated in other cellular pathways involving protein–DNA complexes. Support for this idea comes from Veaute and colleagues' finding² that Srs2 can also remove RecA, an *Escherichia coli* relative of Rad51, from DNA.

In addition, both genetic and physical approaches have shown that yeast cells with mutant Srs2 are defective in certain types of recombination event^{9,12,13}. The implication is that Srs2, as well as preventing an early step in homologous recombination, might actively promote specific recombination subpathways. This apparent paradox could

be resolved if this alternative role of Srs2 is more closely related to its helicase activity, or is controlled by its phosphorylation state, or is influenced by the formation of complexes between Srs2 and other proteins or by the type of DNA damage.

Finally, it remains to be seen how Srs2 itself is regulated. Recombination, of course, is often essential, and so must not always be prevented. ■

Marco Foiani is at the Istituto FIRC di Oncologia Molecolare, Via Adamello 16, 20141 Milano, and the Dipartimento di Genetica e di Biologia dei Microrganismi, Università degli Studi di Milano, Italy.

e-mail: foiani@ifom-firc.it

1. Krejci, L. *et al.* *Nature* **423**, 305–309 (2003).
2. Veaute, X. *et al.* *Nature* **423**, 309–312 (2003).
3. Paques, F. & Haber, J. E. *Microbiol. Mol. Biol. Rev.* **63**, 349–404 (1999).
4. Sogo, J. M., Lopes, M. & Foiani, M. *Science* **297**, 599–602 (2002).
5. Hickson, I. D. *Nature Rev. Cancer* **3**, 169–178 (2003).
6. Rong, L., Palladino, F., Aguilera, A. & Klein, H. L. *Genetics* **127**, 75–85 (1991).
7. Chanet, R., Heude, M., Adouda, A., Maloisel, L. & Fabre, F. *Mol. Cell. Biol.* **16**, 4782–4789 (1996).
8. Liberi, G. *et al.* *EMBO J.* **19**, 5027–5038 (2000).
9. Vaze, M. *et al.* *Mol. Cell* **10**, 373–385 (2002).
10. Gangloff, S., Soustelle, C. & Fabre, F. *Nature Genet.* **25**, 192–194 (2000).
11. Mankoury, H., Craig, T. J. & Morgan, A. *Nucleic Acids Res.* **30**, 1103–1113 (2002).
12. Paques, F. & Haber, J. E. *Mol. Cell. Biol.* **17**, 6765–6771 (1997).
13. Hegde, V. & Klein, H. *Nucleic Acids Res.* **28**, 2779–2783 (2000).

Planetary science

Jupiter's moonopoly

Douglas P. Hamilton

A further 23 satellites have been discovered in orbit around Jupiter. With diameters of between two and eight kilometres, the moons are the smallest yet spotted around any planet.

Swinging serenely around the Sun, mighty Jupiter has reason to be pleased: its pre-eminence as the planet with the largest number of natural satellites, or moons, has been dramatically and decisively re-established. Fending off strong challenges from rival Saturn and wild card Uranus, the Solar System's largest planet now has nearly as many known moons as all of its competitors combined. Satellite-seekers

Scott Sheppard and David Jewitt are responsible for returning Jupiter to its dominant status — on page 261 of this issue¹, they report the discovery of nearly two dozen new jovian moons.

The search for planetary satellites has a long history, dating back to 1610 and Galileo Galilei's discovery of four star-like objects orbiting Jupiter — Io, Europa, Ganymede and Callisto. Saturn's splendid ring system

Planet	Number of satellites	Number of irregular satellites	Largest irregular satellite and radius (km)
Earth	1	—	
Mars	2	—	
Jupiter	60	52	Himalia 85
Saturn	31	14	Phoebe 110
Uranus	22	6	Sycorax 80
Neptune	11	5	Triton 1,353
Pluto	1	—	

Figure 1 Planets and satellites. Irregular, or distant, satellites are found only around the giant planets and are thought to have been captured during the final stages of planetary formation. Total numbers are continually updated at ref. 4 and include this year's findings, up to April 2003: 20 moons at Jupiter, 1 at Saturn and 3 at Neptune. The smallest objects spotted at Jupiter are barely 2 km across¹.

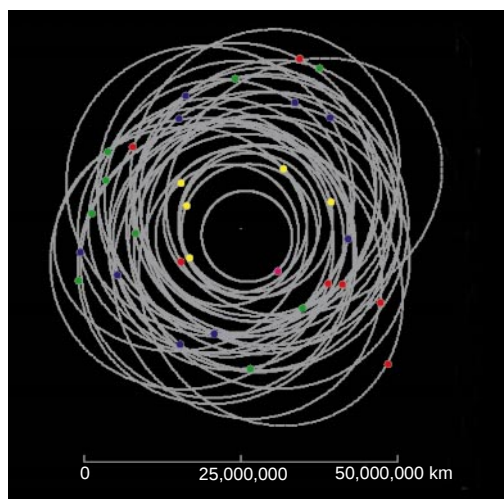


Figure 2 The orbits of Jupiter's outer satellites. The giant planet is at the centre of the image (taken from an animation program¹³), with 31 of its distant satellites grouped into the families identified by Sheppard and Jewitt¹. Prograde objects (pink and yellow) circle Jupiter in the anticlockwise direction when viewed from above Jupiter's north pole, while the three retrograde families (blue, green and red) circle clockwise. The few known prograde orbits nestle inside the far more numerous retrograde ones.

and its largest moon, haze-enveloped Titan, were first seen by Christiaan Huygens about 50 years later, and, in 1684, the discoveries of icy Dione and Tethys established Saturn as the planet with the most moons, a title it held for 230 years. Jupiter's Sinope, spotted in 1914, evened the score at nine known moons apiece, and two additional findings in 1938 allowed the giant planet to surge into the lead. Saturn staged a surprise comeback in 1980, when seven new satellites were spotted by the Voyager spacecraft and ground-based observers. Then came the great upset of 1999: dark horse Uranus revealed three additional outer satellites² and vaulted to the forefront. But the title has since been reclaimed — first by Saturn, with a dozen new discoveries reported³ in 2000, and now by Jupiter, with the 23 findings detailed in this issue¹.

Currently, the number of known planetary moons stands⁴ at 128 (Fig. 1). More than half of this total has been added since 1997, when Brett Gladman and colleagues found the first two distant, or 'irregular', satellites of Uranus⁵. The large number of satellite discoveries over the past six years, at an ever quickening pace, is reminiscent of the situation following Jewitt and Luu's 1992 discovery⁶ of the first trans-neptunian (or Kuiper belt) objects. Both population explosions have been fuelled by major improvements in digital-camera technology⁷.

Nearly two-thirds of the known moons (including all of the recent discoveries) are irregular satellites, orbiting far from their planets along highly tilted, elliptical paths. These objects are believed to have been captured by their planets from independent orbits around the Sun early in the history of the Solar System. Regular satellites, by contrast, have much smaller, untilted, circular orbits, and were probably formed out of the disks of gas and dust that surrounded the giant planets in their youth. Energy dissipation in these early accretion disks also

acted to facilitate the capture of the irregular satellites⁸.

The orbital distributions of irregular satellites at Jupiter (Fig. 2) and Saturn, particularly their tilt angles, show several intriguing patterns that hint at past dynamical and collisional processes. No moons have yet been found on orbits tipped by more than about 55° to the planet's orbital plane. This is due to gravitational forcing from so-called solar tides, which is the largest perturber of distant planetary satellites^{9–11}. Objects with greater tilts experience large, correlated changes in their inclinations and eccentricities which put them on orbits that penetrate deep into the inner jovian system. Such objects are lost as a result of destabilizing gravitational kicks from large regular satellites and, in some cases, through direct collisions¹¹.

The lack of highly tilted orbits means that the orbits may be cleanly divided into prograde (those that circulate in the same direction as the planet's motion around the Sun), and retrograde (those that move in the opposite direction). At Saturn and Neptune, the number of distant, prograde moons is comparable to the number of retrograde objects, but at Jupiter and Uranus, the retrogrades dominate. Also in this issue (page 264), Astakhov and colleagues¹² proffer an explanation for the striking Jupiter–Saturn dichotomy. Through numerical integrations of the capture process, they find that interactions with Jupiter's large regular satellite Callisto preferentially remove moons with prograde orbits, which penetrate closer to Jupiter than retrogrades (see, for example, ref. 10). Titan, closer to Saturn than Callisto is to Jupiter, is less efficient at removing the prograde population. Although alternative explanations for the dichotomy cannot be ruled out, the orbital properties of the extant satellites certainly put important constraints on conditions that existed during the capture process¹².

The prograde and retrograde groups can be further divided into families of objects that share similar characteristics, such as the

size, shape and tilt of their orbits. At least five such families are evident^{1,11} at Jupiter and four are seen³ at Saturn. Like asteroid families, each satellite family appears to have been formed from the collisional breakup of a larger object^{1,3,11}.

So satellites are occasionally shattered by breakups, but do moons ever merge? Although retrograde satellites at Jupiter now vastly outnumber the prograde irregulars, the latter (primarily Himalia) account for nearly 98% of the mass of the distant jovian satellites. And at Saturn, despite similar numbers of prograde and retrograde irregulars, retrograde Phoebe dominates with approximately 99.5% of the orbiting mass. Perhaps Himalia and Phoebe are merely the largest surviving members of much bigger primordial populations. Alternatively, these two large satellites may have grown significantly by cannibalizing some of their smaller neighbours, or by accreting inwardly migrating solids in the latter stages of giant-planet growth. This hypothesis is supported by the fact that both Phoebe and Himalia lie near the inner edge of the zone of irregular satellites (where collisional accumulation rates are fastest) and have relatively low orbital eccentricities and tilts. These and other satellite evolutionary processes need to be explored more fully.

Although ever smaller moons around the giant planets will undoubtedly be discovered, it is likely that Jupiter, whose satellites appear brightest because of their proximity to the Earth and the Sun, will continue to dominate the moon count for the foreseeable future. ■

Douglas P. Hamilton is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742, USA.

e-mail: hamilton@astro.umd.edu

1. Sheppard, S. S. & Jewitt, D. C. *Nature* **423**, 261–263 (2003).

2. Gladman, B. J. et al. *Icarus* **147**, 320–324 (2000).

3. Gladman, B. J. et al. *Nature* **412**, 163–166 (2001).

4. http://ssd.jpl.nasa.gov/sat_discovery.html

5. Gladman, B. J. et al. *Nature* **392**, 897–899 (1998).

6. Jewitt, D. C. & Luu, J. *Nature* **392**, 730–732 (1993).

7. Hamilton, D. P. *Nature* **412**, 132–133 (2001).

8. Pollack, J. B., Burns, J. A. & Tauber, M. E. *Icarus* **37**, 587–611 (1979).

9. Kozai, Y. *Astron. J.* **102**, 1213–1225 (1962).

10. Hamilton, D. P. & Burns, J. A. *Icarus* **92**, 118–131 (1991).

11. Carruba, V. et al. *Icarus* **158**, 434–449 (2002).

12. Astakhov, S. A., Burbanks, A. D., Wiggins, S. & Farrelly, D.

Nature **423**, 264–267 (2003).

13. <http://janus.astro.umd.edu/javadir/orbits/moons.html>

Corrections

● In the News and Views in Brief item "Cell biology: Muscle mice" (*Nature* **422**, 393; 2003), the adult stem cells referred to were derived not from human bone marrow but from synovial membrane.

● The correctly spelt name of the author whose work was discussed by Ian Stewart in "Mathematics: Regime change in meteorology" (*Nature* **422**, 571–573; 2003) is Daan Crommelin.

Obituary

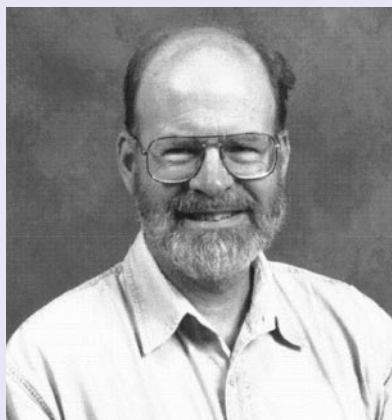
Charles A. Janeway Jr (1943–2003)

Charles A. Janeway Jr, professor of immunobiology at Yale University School of Medicine in New Haven and one of the pre-eminent modern immunologists, died on 12 April 2003 of lymphoma. Most scientists only dream of contributing to a paradigm shift — Janeway personally initiated one. Most focus their research efforts in one narrow area of specialization — Janeway's work ranged over the breadth of immunology. Most influence the training of only a small cohort of students, but Janeway's contribution to education reached thousands, at all levels of sophistication.

In 1989, Janeway launched the idea that, in addition to the ability of the immune system to discriminate between 'self' and 'non-self' molecules, there is a second law of immunology. He reasoned that there must be mechanisms to ensure that an immune response is not only specific, but is also appropriate. In other words, the immune system should respond vigorously and effectively to antigens associated with potentially pathogenic microorganisms, and ignore or devote little energy to fundamentally 'bland' antigens, even if they are foreign. In a seminal paper, Janeway postulated that the immune system would 'know' what is appropriate because of cues provided by the microorganism bearing the antigen, and he introduced the terms 'pathogen-associated molecular patterns' for these cues, and 'pattern-recognition receptors' for the postulated elements on immune cells that would sense the microbial signals.

Now, it is one thing to have a great idea and another to prove it. With Ruslan Medzhitov, a postdoctoral fellow and then colleague at Yale, Janeway showed that mammalian relatives of the Toll protein — a key player in the inflammatory response in fruitflies — constitute one set of pattern-recognition receptors. Indeed, we now know that the Toll-like receptors recognize a wide array of structures associated with various pathogens, including lipopolysaccharide (found in the cell wall of many bacteria) and double-stranded RNA. Recognition of these molecules by key components of the innate immune system, notably dendritic cells, sets off intracellular signals that equip the cells to direct the priming and differentiation of cells of the adaptive immune system. This system can then deal optimally with the type of pathogen that induced the response.

Janeway was the product of a distinguished medical family. His great-



Immunologist who postulated a second law of immunology

grandfather had been health commissioner of New York City; his grandfather described 'Janeway's lesions', a key sign of subacute bacterial endocarditis (infection of the inner surface and valves of the heart); and his father was physician-in-chief of the Boston Children's Hospital and one of the founders of the modern study of immunodeficiency diseases. Janeway followed in this tradition professionally, though not sartorially — despite his Brahmin background, he was most comfortable in a workshirt and jeans, wearing his trademark red hat and looking more like a holdover from the flower-power days of the 1960s than the Yale professor he was for so many years.

Janeway's education was critically influenced during his medical school days by Hugh McDevitt, who advised him to work in England with immunologist John Humphrey. Janeway credited the two years he spent with Humphrey — from 1965 to 1967 — as being extraordinarily influential in shaping his life as a scientist. After clinical training, he came to the Laboratory of Immunology at the National Institute of Allergy and Infectious Diseases (NIAID) in 1970, as a postdoctoral fellow with one of us (W.E.P.). Janeway was of that imaginative breed of scientist for whom the right environment and largely being left alone constituted the best training.

After leaving the NIAID, a sojourn at the University of Uppsala was followed by an appointment in 1977 to the pathology department at Yale, and then to the immunobiology section. In New Haven, Janeway married another immunologist, Kim Bottomly, whom he had met at the

NIAID. This relationship was a deep one, encompassing home, family and science. In particular, Janeway credited his wife with encouraging him to articulate his thoughts about the connection between the innate and adaptive immune systems.

While best known for bringing this connection to the fore, Janeway also made substantial contributions to a range of other topics. He championed the notion of the CD4 and CD8 proteins as 'co-receptors' in the physically intimate interaction between MHC molecules (which present antigens to immune cells) and the T-cell antigen receptor. He made advances in our understanding of cell-mediated autoimmune disease, T- and B-cell tolerance of self molecules, the role of co-signalling in T-cell activation, thymocyte selection, and the biochemical basis of T-cell recognition of antigen-MHC complexes. He collaborated with Bottomly in her discovery of two distinct functional subsets of CD4-expressing cells, which she called T inflammatory and T helper cells — now known as Th1 and Th2 cells.

Janeway was a superb teacher, and for many years ran the Yale immunology course for medical students. He used his lecture notes for this course to develop *Immunobiology*, which has become one of the major immunology textbooks. This text was a central part of Janeway's intellectual life, and even while ill he brought out new editions that kept pace with the rapidly changing face of modern immunology. Similarly, although the idea of a link between adaptive and innate immunity came to him before he was diagnosed with lymphoma, virtually all the major research that established the concept was done afterwards. His underlying love of the field — of the wrestling, through experiment, discussion and thought, of each new insight into the immune system and the cells that ultimately turned against him — always showed through.

In 1998, Charlie opened his presidential address to the American Association of Immunologists with a recitation of Robert Frost's "The Road Not Taken". He used it to make the point that "the road less traveled by" was emblematic of the choices one makes that define one's life — in his case, his decision to investigate the link he had postulated between innate and adaptive immunity. For Charlie, those choices led to a legacy that will continue to enrich modern science. William E. Paul and Ronald N. Germain
William E. Paul and Ronald N. Germain are in the Laboratory of Immunology, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA.
e-mail: wpaul@niaid.nih.gov

Immunology

Assisted suicide
for B cells

J. Exp. Med. **197**, 1125–1139 (2003)

Infectious organisms have a variety of ways of protecting themselves from the immune systems of their unwitting hosts. Carl S. Goodyear and Gregg J. Silverman now show that, for the bacterium *Staphylococcus aureus*, attack is the best form of defence.

When the immune system's B cells encounter unfamiliar items, such as material from a bacterium, they normally respond by differentiating into cells that produce antibodies against the pathogen. They also generate long-lasting memory B cells, which deal rapidly with the intruder if reinfection occurs.

Goodyear and Silverman find that these events do not take place when mice are exposed to a protein called SpA, produced by *S. aureus*. The protein binds to B cells and causes them to commit suicide within a few hours. Memory B cells never have the chance to develop. The authors believe that these processes also occur during natural *S. aureus* infections.

The findings could aid the development of vaccines against *S. aureus*. But what Silverman and colleagues are currently working on is a treatment for systemic lupus erythematosus, an autoimmune disease caused by faulty B cells. The idea is to use the bacterial SpA protein to kill the subset of B cells that cause the disease. **Lisen Arnheim**

Microfluidics

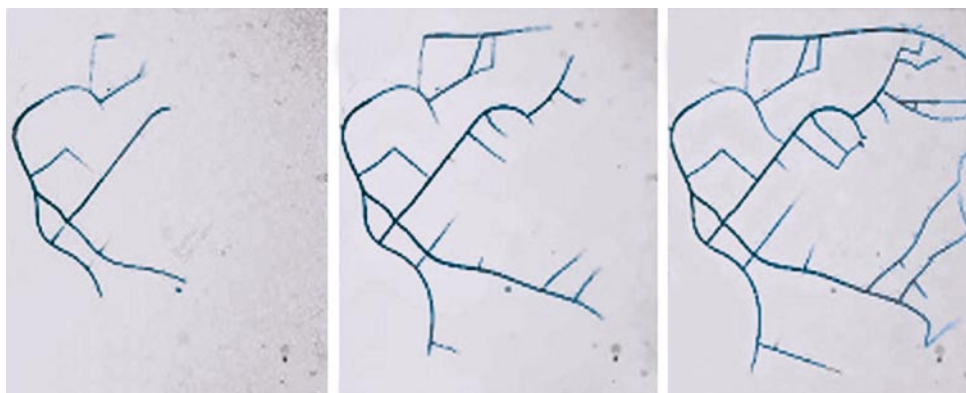
Traffic flow

Langmuir doi:10.1021/la030054x (2003)

What's the fastest route across downtown Boston? Michael J. Fuerstman and colleagues can tell you, by using a microfluidic system to find the way through the complex maze of the city and identify the best path.

Fuerstman *et al.* have built a number of mazes — including one representing a street plan of Boston — from networks of microchannels imprinted in poly(dimethylsiloxane), sealed with a flat surface. The mazes are filled with one fluid, then a contrasting dye solution is introduced and its movements traced as it explores, under an applied pressure gradient, all continuous paths between inlet and outlet.

The inlet of one such system (pictured) marks the site of the Boston Museum of Science, and is connected to the outlet — the Computer Museum — through channels representing some of the city's streets and highways. The different widths of the microfluidic channels control the velocity of the fluid to roughly match the



Which way? Dye running through a microfluidic representation of the streets of Boston signals the quickest route across town. These images were taken over a five-second interval.

relative driving speeds on the corresponding roads. Introducing a blue liquid into the pre-filled network reveals two routes between the museums, one slightly faster than the other.

In a maze possessing multiple solutions, the tracing fluid reaches the outlet fastest along the path of least fluidic resistance. Adjusting the properties of the channels and fluids will affect the compromise found between the shortest and widest channels. The approach, suggest the authors, might be used to model complex nonlinear problems such as traffic flow or distribution systems.

Magdalena Helmer

Cancer

Brake for cell division

Genes Dev. **17**, 1201–1206 (2003)

The *tob* gene is one of many that prevent cells from dividing endlessly. Yutaka Yoshida and colleagues now provide evidence that *tob* also suppresses tumour formation.

It was already known that cells without Tob protein continue to multiply under circumstances where other cells give up — during famine, for instance. But it was not clear whether Tob is connected with the uncontrolled cell growth that leads to tumour formation.

Enter Yoshida *et al.*, who have generated *tobless* mice. The authors find that tumours grow spontaneously in various tissues in these mice, including the lymph nodes and lungs. In addition, more *tob*-deficient animals than control mice develop liver tumours, especially after treatment with a liver-specific cancer-causing compound. This is of interest as there are currently no adequate animal models for liver cancer.

Yoshida *et al.* further show that Tob can repress the expression of cyclin D1, which drives cell proliferation. So, when Tob activity is reduced, cyclin D1 levels rise, and cell division goes into overdrive. Although analysis of more than 50 human

tumours did not reveal any mutations in the *tob* gene, its expression level was reduced in several human lung tumours. So it seems that suppression of Tob expression may contribute to tumour progression.

Marie-Thérèse Heemels

Biomaterials

Size matters

Proc. Natl Acad. Sci. USA **100**, 5597–5600 (2003)

Biological composite materials such as bone, shell and tooth are renowned for their strength. They generally consist of crystals or platelets of a hard mineral, often calcium carbonate, embedded in a soft matrix of proteins and other organic molecules. Huajian Gao and colleagues say that the size of the mineral particles — typically tens to hundreds of nanometres — is not arbitrary. Rather, it seems to be tuned to give the mineral a remarkable characteristic: it is insensitive to the presence of structural flaws.

In hard, brittle materials, flaws such as microscopic scratches or holes are an Achilles' heel, with a disruptive potential that belies their size. Stresses in the solid become concentrated at these places, nucleating cracks that travel rapidly through the material. But the analysis of Gao *et al.* shows that the fracture resistance of a plate containing a surface flaw increases as the plate thickness decreases, becoming equal to that of a perfect, unflawed plate at a thickness of around 30 nm. The exact value depends on the precise properties of the material, but it seems clear that the weakening effect of flaws vanishes specifically at the nanoscale.

The structure of these biomineral particles is also probably governed by constraints on, for example, their nucleation and growth rates. But it seems that attempts to recreate the superior properties of biomaterials in synthetic composites will require mimicry not just of the arrangements of the constituent materials but also of their scale. **Philip Ball**

Mission to Earth's core — a modest proposal

Not science fiction, but a technically feasible plan to probe our planet's inner workings.

Planetary missions have enhanced our understanding of the Solar System and how planets work, but no comparable exploratory effort has been directed towards the Earth's interior, where equally fascinating scientific issues are waiting to be investigated. Here I propose a scheme for a mission to the Earth's core, in which a small communication probe would be conveyed in a huge volume of liquid-iron alloy migrating down to the core along a crack that is propagating under the action of gravity. The grapefruit-sized probe would transmit its findings back to the surface using high-frequency seismic waves sensed by a ground-coupled wave detector. The probe should take about a week to reach the core, and the minimum mass of molten iron required would be 10^8 – 10^{10} kg — or roughly between an hour and a week of Earth's total iron-foundry production.

We live on the Earth's surface, which divides what is above from what is below (Fig. 1). The part above us (the rest of the Universe) is mostly empty, mostly unknown and about 10^{57} times larger by volume. The part below is crammed with interesting stuff and is also mostly unknown, despite its much greater proximity to us. Space probes have so far reached a distance of about 40 astronomical units (6×10^9 km), but subterranean probes (drill holes) have descended only some 10 km into the Earth.

Travel downwards is impeded by the dense intervening matter, and the energy required to penetrate it by melting is about 10^9 times (per unit distance travelled) the energy needed for space travel — a fact that partly explains the large difference in distances attained¹. Travel downwards has also been impeded by the much more limited allocation of financial and material resources relative to those provided for space travel — there is no underground equivalent of NASA.

One possible means of reaching the core appeals to the 'China syndrome' idea² and requires melting of the rock, but the trip times in these scenarios are thousands of years or more — geologically short but too long on a human timescale for any government to contemplate funding. However, a liquid-iron-filled crack initiated in the Earth would propagate downwards (despite very high pressures), closing up behind as it travels, and a neutrally buoyant, insoluble probe could be carried along for the ride.

I use well-established principles of magma fracturing^{3,4} (the migration of melt through the Earth's lithosphere). Consider a vertical crack of approximate width d . The other horizontal dimension is $W \gg d$. Let the characteristic vertical extent be L . Provided that

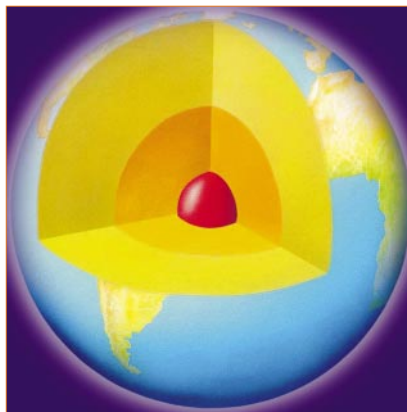


Figure 1 Next to the riches lavished on space exploration, an unmanned journey to the centre of the Earth looks almost frugal.

L is sufficiently large, the propagation speed will be limited by the channel-flow velocity of the fluid. The relevant solution has turbulent flow, with the crack-propagation speed, V_{prop} , being roughly equal to the channel-flow velocity, or $[(\Delta\rho/\rho)gd^{5/4}/\nu^{1/4}]^{4/7}$, where $\Delta\rho \approx \rho$ is the density difference between melt and surrounding rock, g is the gravitational acceleration, and ν is the kinematic viscosity of liquid iron (about 10^{-6} m² s⁻¹). Thus, $V_{\text{prop}} \approx 30d^{5/7}$ m s⁻¹, with d in metres.

A crack of width d and length L would have associated deviatoric stresses of about $\mu d/L$, where $\mu \approx 10^{11}$ pascals (Pa) and is the shear modulus of the rock. This stress will be overcome by the pressure head in the crack when it is comparable to $\Delta\rho gL$. This defines a natural crack of length $L \approx (\mu d/\Delta\rho g)^{1/2} \approx (1 \text{ km}) \times d^{1/2}$ and with a stress level of about $3 \times 10^7 d^{1/2}$ Pa. If stresses of about 10^7 Pa are sufficient (as studies of lithospheric cracking suggest), $d \approx 0.1$ m and $L \approx 300$ m. Assuming that the other dimension, W , is also around 300 m, the volume of iron contained in the crack will be about 10^4 m³, or 10^8 kg, which is the amount produced in about an hour by the world's foundries⁵.

As the crack propagates downwards at about 5 m s⁻¹ (a speed that would give a mission timescale of around a week), the released gravitational energy would cause heating and partial melting of the silicate rock walls¹. The work required to initiate the crack is plausibly about μLWd , or about 10^{15} joules, which is equivalent to a few megatons of TNT explosive, an earthquake of magnitude 7 on the Richter scale, or a nuclear device with a capability that is within the range of those currently stockpiled. Some aspects of crack migration in the deep Earth may involve modified, less optimistic values¹, however, requiring an iron volume that is

greater by one or two orders of magnitude.

It may also be feasible to make use of existing favourable stress environments in the Earth and to avoid the use of nuclear devices. The technological challenge of initiating the crack should be less than that posed by the Manhattan Project.

The embedded, solid-state probe could plausibly have a volume of d^3 , or roughly that of a grapefruit. It would be a high-melting-point alloy in saturation equilibrium with the neighbouring liquid-iron alloy and would contain miniaturized instrumentation for measuring temperature and electrical conductivity, detecting abundant and trace elements, and so on — details that would require an instrument-development programme. The Earth's interior is opaque to electromagnetic signals with periods of less than the mission timescale, and neutrinos are difficult to use, so acoustic communication would be best.

I assume a probe power, P , of about 10 watts throughout the mission¹ (similar to that of some current deep-space missions) and treat the probe as a monopole source of compressional acoustic radiation. Let the amplitude of the oscillatory motion of the probe surface be x . For angular frequency ω and wave-propagation speed c , $P = 4\pi\rho x^2\omega^4 d^4/c$, and the far-field wave-displacement amplitude would be about $\omega d^2 x/rc$, where r is the distance from the source⁶. For $f \equiv \omega/2\pi$ and $c \approx 10^4$ m s⁻¹, $x \approx 300(10^2 \text{ Hz}/f)^2 \mu\text{m}$ and the amplitude at the Earth's surface would be about $10^{-13}(10^2 \text{ Hz}/f)$ m, roughly the radius of an atomic nucleus.

For the high frequencies of interest here, the quality factor of mantle rock⁷ is about 10^4 , but frequencies in excess of 100 Hz would nonetheless be unacceptably attenuated. However, frequencies much smaller than this would be a problem because of natural seismic energy, part of the reason that the Laser Interferometer Gravitational-wave Observatory (LIGO; a device that seeks to detect gravitational radiation)⁸ operates in the kilohertz range. The encoding of the signal would greatly aid its detection, and the amplitude of the signal would be within the detectability limit of LIGO (were it coupled to the ground, rather than being decoupled as at present). For the duration of the mission, about 10^8 cycles of probe oscillation would occur, which would be sufficient to encode the state and composition of the deep Earth.

This proposal is modest compared with the space programme, and may seem unrealistic only because little effort has been devoted to it. The time has come for action.

David J. Stevenson

California Institute of Technology, Pasadena,

California 91125, USA
e-mail: djs@gps.caltech.edu

1. <http://www.gps.caltech.edu/faculty/stevenson/coremission>
2. Turcotte, D. L. & Emmerman, S. H. *J. Geophys. Res.* **88**, B91–B96 (1983).
3. Schubert, G., Turcotte, D. L. & Olson, P. *Mantle Convection in*

- the Earth and Planets* 166–176 (Cambridge Univ. Press, 2001).
4. Weertman, J. *J. Geophys. Res.* **76**, 1171–1183 (1971).
 5. <http://www.worldsteel.org>
 6. Landau, L. D. & Lifshitz, E. M. *Fluid Mechanics* (Pergamon, London, 1959).
 7. Anderson, D. L. & Given, J. *J. Geophys. Res.* **87**, 3893–3904 (1982).
 8. <http://www.ligo.caltech.edu>

Competing financial interests: declared none.

Aetiology

Koch's postulates fulfilled for SARS virus

Severe acute respiratory syndrome (SARS) has recently emerged as a new human disease, resulting globally in 435 deaths from 6,234 probable cases (as of 3 May 2003). Here we provide proof from experimental infection of cynomolgus macaques (*Macaca fascicularis*) that the newly discovered SARS-associated coronavirus (SCV) is the aetiological agent of this disease. Our understanding of the aetiology of SARS will expedite the development of diagnostic tests, antiviral therapies and vaccines, and may allow a more concise case definition for this emerging disease.

According to Koch's postulates, as modified by Rivers for viral diseases, six criteria are required to establish a virus as the cause of a disease¹. The first three criteria — isolation of virus from diseased hosts, cultivation in host cells, and proof of filterability — have been met for SCV by several groups^{2–5}. Moreover, of 96 individuals complying with

the World Health Organization's definition of SARS⁶ in Hong Kong, 86 (90%) yielded laboratory evidence of SCV infection.

We have tested for the three remaining criteria: production of comparable disease in the original host species or a related one, re-isolation of the virus, and detection of a specific immune response to the virus. We inoculated two macaques with Vero-cell-cultured SCV isolated from a fatal SARS case, and monitored their clinical signs, virus excretion and antibody response. The animals were killed six days post-inoculation (d.p.i.), and we then carried out gross and histopathological examinations of them.

Both SCV-inoculated macaques became lethargic from 3 d.p.i. onwards and developed a temporary skin rash, and one suffered respiratory distress from 4 d.p.i. onwards. The macaques excreted virus from the nose and throat at 2–6 d.p.i., as shown by polymerase chain reaction with reverse transcription (RT-PCR) and by virus isolation (see supplementary information). The isolated virus was identical to that inoculated, as shown by negative-contrast electron microscopy (Fig. 1a) and RT-PCR analysis. Seroconversion to

SCV, as determined by indirect immunofluorescence assay using infected Vero cells, was demonstrated in two other SCV-infected macaques at 16 d.p.i.. The virus was also isolated from the faeces of one of these animals (see supplementary information).

At gross necropsy, one macaque had severe multifocal pulmonary consolidation, and SCV infection was detected in lung tissue by RT-PCR and virus isolation. Histologically, both macaques had interstitial pneumonia of differing severity. The one with gross lesions had diffuse alveolar damage, marked by necrosis of alveolar and bronchiolar epithelium and flooding of alveolar lumina with proteinaceous fluid, admixed with fibrin, erythrocytes, alveolar macrophages and neutrophils (Fig. 1b). Occasional multinucleated cells (syncytia) were present in the lumen of bronchioles and alveoli (Fig. 1c). These lesions are indistinguishable from those in biopsied lung tissue and in autopsy material from SARS patients⁵, including the presence of syncytia in alveolar lumina⁴.

SCV thus fulfils all of Koch's postulates as the primary aetiological agent of SARS. This does not exclude the possibility that other pathogens, including human metapneumovirus (hMPV) and *Chlamydia pneumoniae*, may have exacerbated the disease in some SARS patients. However, these were not present in SCV-inoculated macaques (results not shown), were not found consistently in SARS patients, and do not usually cause the lesions associated with SARS. Moreover, lesions in macaques infected experimentally with hMPV isolated from a non-SARS individual⁷ were limited to mild suppurative rhinitis and minimal erosion in conducting airways, and disease was not exacerbated in two SCV-infected macaques subsequently inoculated with hMPV (results not shown).

Ron A. M. Fouchier*, **Thijs Kuiken***, **Martin Schutten***, **Geert van Amerongen***, **Gerard J. J. van Doornum***, **Bernadette G. van den Hoogen***, **Malik Peiris†**, **Wilina Lim‡**, **Klaus Stöhr§**, **Albert D. M. E. Osterhaus***

*Department of Virology, Erasmus Medical Centre, 3015 GE Rotterdam, The Netherlands

e-mail: a.osterhaus@erasmusmc.nl

†Department of Microbiology, University of Hong Kong, University Pathology Building, Queen Mary Hospital, Hong Kong SAR, China

‡Government Virus Unit, 9/F Public Health Laboratory Centre, 382 Nam Cheong Street, Shek Kip Mei, Kowloon, Hong Kong SAR, China

§On behalf of members of the SARS Aetiology Study Group, World Health Organization, Avenue Appia 20, CH-1211, Geneva 27, Switzerland

1. Rivers, T. M. *J. Bacteriol.* **33**, 1–12 (1937).

2. Poutanen, S. M. et al. *N. Engl. J. Med.* (in the press).

3. Drosten, C. et al. *N. Engl. J. Med.* (in the press).

4. Ksiazek, T. G. et al. *N. Engl. J. Med.* (in the press).

5. Peiris, J. S. M. et al. *Lancet* **361**, 1319–1325 (2003).

6. *Wkly Epidemiol. Rec.* **78**, 81–83 (2003).

7. van den Hoogen, B. G. et al. *Nature Med.* **7**, 719–724 (2001).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

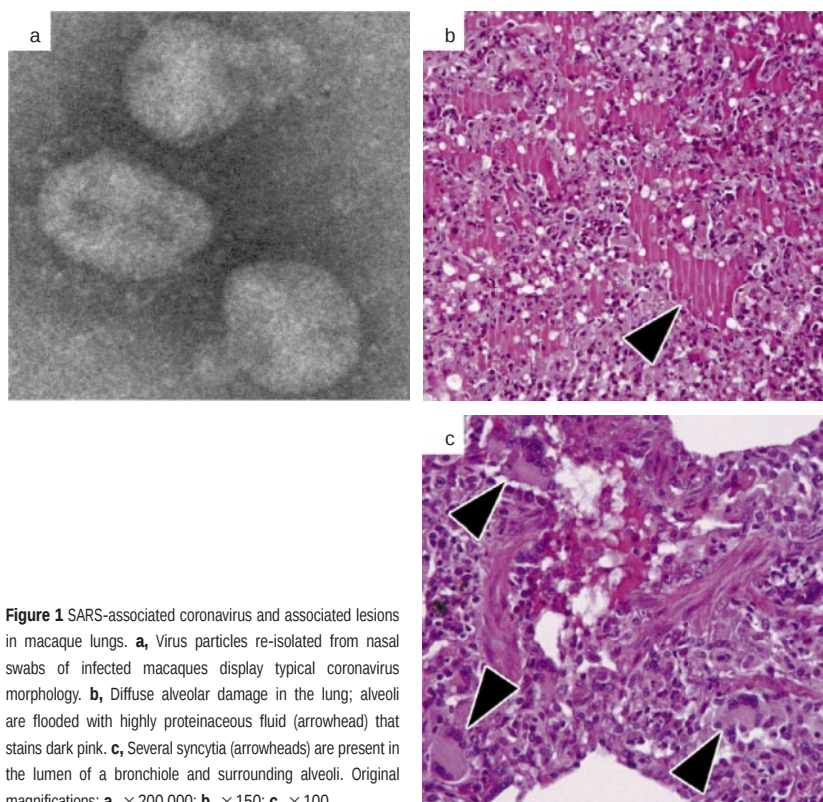


Figure 1 SARS-associated coronavirus and associated lesions in macaque lungs. **a**, Virus particles re-isolated from nasal swabs of infected macaques display typical coronavirus morphology. **b**, Diffuse alveolar damage in the lung; alveoli are flooded with highly proteinaceous fluid (arrowhead) that stains dark pink. **c**, Several syncytia (arrowheads) are present in the lumen of a bronchiole and surrounding alveoli. Original magnifications: **a**, $\times 200,000$; **b**, $\times 150$; **c**, $\times 100$.

Sequencing and comparison of yeast species to identify genes and regulatory elements

Manolis Kellis^{*†}, Nick Patterson^{*}, Matthew Endrizzi^{*}, Bruce Birren^{*} & Eric S. Lander^{*‡}

^{*} Whitehead/MIT Center for Genome Research, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

[†] Department of Computer Science and [‡] Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

Identifying the functional elements encoded in a genome is one of the principal challenges in modern biology. Comparative genomics should offer a powerful, general approach. Here, we present a comparative analysis of the yeast *Saccharomyces cerevisiae* based on high-quality draft sequences of three related species (*S. paradoxus*, *S. mikatae* and *S. bayanus*). We first aligned the genomes and characterized their evolution, defining the regions and mechanisms of change. We then developed methods for direct identification of genes and regulatory motifs. The gene analysis yielded a major revision to the yeast gene catalogue, affecting approximately 15% of all genes and reducing the total count by about 500 genes. The motif analysis automatically identified 72 genome-wide elements, including most known regulatory motifs and numerous new motifs. We inferred a putative function for most of these motifs, and provided insights into their combinatorial interactions. The results have implications for genome analysis of diverse organisms, including the human.

Extracting the complete functional information encoded in a genome—including the genic, regulatory and structural elements—is a central challenge in biological research. Ideally, one would be able to extract this information directly from the DNA sequence itself without recourse to extensive experimentation. At present, however, our ability to directly interpret genomes is rudimentary.

De novo identification of the complete set of protein-coding sequences remains imperfect, even in well-studied organisms with compact genomes. The yeast *Saccharomyces cerevisiae*, for example, has enjoyed a complete genome sequence since 1996 (ref. 1). However, the number of biologically significant open reading frames (ORFs) has been the subject of considerable debate^{2–6}—with proposed counts ranging from roughly 4,800 to 6,400. The situation is worse for organisms with large, complex genomes, such as mammals. Because the results of *de novo* gene prediction are not sufficiently reliable, gene identification is typically based on a comparison of genomic sequence to known messenger RNA transcripts or to genes in other organisms⁷.

Direct identification of the repertoire of regulatory elements in a genome is even more challenging. The best approach so far has relied on clustering genes into functionally related subsets (for example, genes with a common biochemical function or coordinated transcription) and then searching for common sequence motifs in the general vicinity of the genes (for example, using computer programs such as MEME⁸ or AlignACE⁹; reviewed in ref. 10). However, the approach has notable drawbacks. It requires extensive prior knowledge about gene function, which is often not available and never comprehensive, and is inherently limited in its ability to extract information from a single genome.

Comparative genome analysis of related species should provide a powerful and general approach for identifying functional elements without previous knowledge of function. Because evolution relentlessly tinkers with genome sequence and tests the results by natural selection, such elements should stand out by virtue of having a greater degree of conservation across related species. The approach has the advantage that one can increase its power by increasing the number of species studied.

Recent studies have demonstrated the potential power of comparative genomic comparison. Cross-species conservation has

previously been used to identify putative genes or regulatory elements in small genomic regions^{11–14}. Light sampling of whole-genome sequence has been studied as a way to improve genome annotation^{5,15}. Complete microbial genomes have been compared to identify pathogenic and other genes^{16–19}. Genome-wide comparison has been used to estimate the proportion of the mammalian genome under selection⁷.

The goal of this paper is to develop and apply general approaches for systematic analysis of protein-coding and regulatory elements within any genome by means of whole-genome comparisons with several related species. We focused on *S. cerevisiae*, because it is the best-studied eukaryote and thus provides the best setting to test the approach. High-quality draft genome sequences from three related *Saccharomyces* species were produced, aligned and analysed to identify protein-coding genes and regulatory elements.

The comparative analysis indicates that *S. cerevisiae* contains only 5,538 genes encoding ≥ 100 amino acids. It proposes the elimination of about 500 previously annotated ORFs; the merger of 33 pairs of consecutive genes; the redefinition of start or stop codons in at least ~ 300 cases; and the existence of ~ 60 new introns. It also identifies 188 genes encoding between 50 and 99 amino acids, including 43 that had previously escaped notice.

Analysis of intergenic regions reveals 72 well-conserved sequence motifs that occur frequently throughout the yeast genome. The set includes most known regulatory motifs, together with a comparable number of previously uncharacterized motifs. We also identified combinatorial relationships among these motifs.

The results demonstrate that comparative genomic analysis of multiple related species has substantial power to identify key functional elements without previous biological knowledge.

Genome structure and evolution

Genome sequencing

Saccharomyces paradoxus, *S. mikatae* and *S. bayanus*, which are separated from *S. cerevisiae* by an estimated 5–20 million years of evolution, were selected for sequencing on the basis of their phylogeny. The three species were found to have sufficient sequence

similarity to *S. cerevisiae* to allow orthologous regions to be aligned reliably, but sufficient sequence divergence to allow many functional elements to be recognized by their greater degree of conservation by a four-way species comparison. All three are members of the *Saccharomyces sensu stricto* group.

For each species, we generated about sevenfold redundant coverage in paired end sequences from a whole-genome shotgun plasmid library (see Methods). The information was then assembled with the Arachne computer program²⁰ into contigs (continuous blocks of uninterrupted sequence) and scaffolds (contigs linked by paired forward–reverse reads from the same plasmid) to yield a draft sequence (Table 1).

The draft genome sequence of each species has long-range continuity (N50 scaffold length of 230–500 kilobases (kb), as compared with 942 kb for the finished sequence of *S. cerevisiae*), relatively short sequence gaps (0.6–0.8 kb, which is small compared with a typical gene), and contains most of the genome (~95%). The data are freely available through public sequence databases and through the *Saccharomyces* Genome Database (SGD) maintained at Stanford (<http://genome-www.stanford.edu/Saccharomyces/>).

Genome alignment

We sought to align the *S. cerevisiae* genome with that of each of the other three species. The first step was to produce a large-scale alignment of genomic regions. Each ORF in *S. cerevisiae* (a total of 6,235 ORFs in the current SGD annotation; see below) was analysed relative to each of the three related species to determine whether it had a clear one-to-one match, multiple ambiguous matches or no match (see Methods). The one-to-one matches were used as orthologous landmarks to define the large-scale alignment, which showed strong conservation of synteny across the species (see below). The second step involved generating local nucleotide-level alignments around each orthologous ORF (see Methods).

Genome evolution at large scale

Most ORFs have clear one-to-one orthologous matches in each species, providing a dense set of landmarks (average spacing ~2 kb) to define blocks of conserved synteny covering essentially the entire genome (Fig. 1). For a small number of ORFs (211), the correspondence is ambiguous, however. These ambiguous matches almost all reflect local gene-family expansion or contraction.

Most of the ambiguities are markedly clustered in telomeric regions (Fig. 2). More than 80% fall into one of 32 clusters of two or more genes (average size ~18 kb, together comprising ~4% of the genome), which correspond nearly perfectly to the 32 telomeric regions of the 16 chromosomes of *S. cerevisiae*. Only one telomeric region lacks a cluster (chromosome VI-R). And only one cluster does not lie in telomeric regions in *S. cerevisiae*: it is a recent insertion of a segment that is telomeric in the other three species.

The rapid structural evolution in the telomeric regions can also

be observed at the gene level. The gene families contained within these regions (including the HXT, FLO, PAU, COS, THI and YRF families) show significant changes in number, order and orientation. The regions also harbour many new sequences, including protein-coding sequences, in the four yeasts (see below). Moreover, the telomeric regions have undergone 11 reciprocal translocations across the species.

Together, these features define relatively clear boundaries for the telomeric regions on all 32 chromosome arms, with sizes ranging from ~7 kb to ~52 kb on chromosome I-R (Fig. 2). The extraordinary genomic churning occurring in these regions probably has a key role in rapidly creating phenotypic diversity over evolutionary time. A high degree of variation in telomeric gene families has also been reported in *Plasmodium falciparum*²¹, the parasite responsible for malaria, and is related to antigenic variation.

Outside of the telomeric regions, few genomic rearrangements are found relative to *S. cerevisiae* (see Methods). We enumerated all non-telomeric rearrangements affecting multiple consecutive genes (Fig. 2). *Saccharomyces paradoxus* shows no reciprocal translocations, four inversions and three segmental duplications; *S. mikatae* shows four reciprocal translocations and 13 inversions; *S. bayanus* has five reciprocal translocations and three inversions. The results confirmed four recently reported reciprocal translocations in these species, identified by pulsed-field gel electrophoresis²², and identified four additional reciprocal translocations that had been missed.

The sequence at the chromosomal breakpoints suggested the possible mechanism that underlies the rearrangements. Notably, the 20 inversions are all flanked by transfer RNA genes in opposite transcriptional orientation and usually of the same isoacceptor type; the origins of inversions in recombination between tRNA genes has not been noted previously. The reciprocal translocations occurred between Ty elements in seven cases and between highly similar pairs of ribosomal protein genes in two cases; the implication of Ty elements in reciprocal translocation is consistent with previous reports^{22–25}. One segmental duplication involves ‘donor’ and ‘recipient’ regions that are descendants of an ancient duplication in the yeast genome²⁶. Differential gene loss of anciently duplicated genes has been reported previously²⁷, but this is the first observation of a recent re-duplication event within anciently duplicated regions.

Genome evolution at the nucleotide level

With sequence alignments at millions of positions across the four species, it is possible to obtain a precise estimate of the rate of evolutionary change, including substitutions and insertions or deletions (indels), in the phylogenetic tree connecting the species (Fig. 3; see also Supplementary Information and <http://www-genome.wi.mit.edu/seq/Saccharomyces/>). Using *S. bayanus* as an outgroup, the substitution rate is similar in *S. cerevisiae* and *S. mikatae*, but is ~67% lower in the lineage leading to *S. paradoxus*. We compared the rate of sequence change at aligned sites across the four species in intergenic and genic (defined here as protein-coding) regions. The proportion of sites corresponding to a different nucleotide in at least one of the three species is 58% in intergenic regions but only 30% in genic regions—a difference of about twofold. The proportion of sites corresponding to an insertion or deletion (indel) is 14% in intergenic regions, but only 1.3% in genic regions. The contrast is even sharper when one considers only indels with length not a multiple of three (which would disrupt a reading frame): 10.2% in intergenic regions versus 0.14% in genic regions—a difference of about 75-fold.

The four genomes show tremendous conservation of synteny and they can be well aligned at the nucleotide level. The overall rate of sequence divergence across the species seems to be high enough to facilitate recognition of functional elements, and the marked contrast between intergenic and genic regions should allow for greatly improved gene identification.

Table 1 Genome assemblies of four *Saccharomyces* species

	<i>S. cerevisiae</i>	<i>S. paradoxus</i>	<i>S. mikatae</i>	<i>S. bayanus</i>
Sequence coverage (fold)	Finished	7.7	5.9	6.4
Genome sequence in contigs (Mb)*	12.16	11.57	11.22	11.32
Genome length, including gaps (Mb)†	12.16	11.75	12.12	11.54
Percentage of genome in contigs	100	98	93	98
Number of scaffolds	16	51	90	100
N50 scaffold length (kb)	942	509	334	234
N50 contig length (kb)	942	51	20	25
Gaps per 100 kb	0	3.2	10.3	4.4
Average gap length (bp)	0	583	847	679

* Sum of contig lengths.

† Sum of scaffold lengths, including contig lengths and estimated gap sizes between consecutive contigs.

Identification of genes

De novo identification of protein-coding genes from genomic sequence is extremely important, yet surprisingly challenging. The basic approach is to identify ORFs that are too long to have occurred by chance. However, stop codons occur at a frequency of only about 1 in 20 in random sequence. Thus, ORFs of ≥ 60 amino acids will occur frequently by chance ($\sim 5\%$ under a simple Poisson model) and even ORFs of ≥ 150 amino acids (or 450 base pairs (bp)) will appear by chance in a large genome (approximately 0.05%). This poses a huge challenge for higher eukaryotes, in which genes are typically broken into many small exons (typical size ~ 125 bp for internal coding exons in mammals²⁸). Yeast genes typically contain fewer and larger exons, but nonetheless there are many true exons encoding ≤ 150 amino acids and it can be difficult to discriminate these from randomly occurring ORFs.

Comparative genomic analysis provides a simple solution: test the ORFs seen in one species by observing whether the orthologous sequence in related species also encodes an ORF. True protein-coding ORFs will typically be under strong selective pressure to preserve the open reading frame, whereas spurious ORFs will accumulate frameshifts and stop codons. We applied this idea to perform *de novo* gene identification in *S. cerevisiae*. We started with the list of all ORFs greater than a given size, compared the list to the three related species and generated a yeast gene catalogue.

Public yeast gene catalogue

When the yeast genome sequence was completed¹, the authors identified 6,275 ORFs in the nuclear genome that could theoretically encode proteins of ≥ 100 amino acids and that do not overlap a longer ORF by more than half of their length. SGD has since updated the catalogue on the basis of complete resequencing and re-annotation of chromosome III, re-analysis of other chromosomes and reports in the scientific literature. This resulted in a current version (as of May 2002) with 6,062 ORFs encoding ≥ 100 amino acids, consisting of 3,966 'named' genes (described in at least one publication) and 2,096 'uncharacterized' ORFs. SGD also includes a small collection of ORFs encoding < 100 amino acids (see below).

Reading frame conservation test

We developed a reading frame conservation (RFC) test to classify each ORF in *S. cerevisiae* as biologically meaningful or meaningless,

on the basis of the proportion of the ORF over which reading frame is locally conserved in each of the other three species (see Methods). For overlapping ORFs in the *S. cerevisiae* genome ($n = 948$), the RFC was calculated only for the portion unique to each overlapping ORF. For spliced genes ($n = 240$), the RFC was calculated only on the largest exon.

Figure 4 illustrates the case of an ORF of 333 bp that is clearly biologically meaningless. The orthologous sequence in all four species is laden with frameshifts (as well as stop codons).

The RFC scores show a clearly bimodal distribution for each of the three species, allowing a decision threshold to be defined for each species to accept an ORF as biologically meaningful or reject an ORF as spurious; that is, occurring by chance. Each species thus 'votes' on the validity of each ORF (or 'abstains' due to lack of orthologous sequence resulting from either incomplete sequence coverage or genomic rearrangement in the species). The votes from the species are then tallied to reach a decision for each ORF (see Methods).

To investigate the power of the approach to reject spurious ORFs, we applied it to a set of control sequences consisting of 340 intergenic sequences in *S. cerevisiae* with lengths similar to the ORFs tested. About 96% were rejected as having conservation properties incompatible with a biologically meaningful ORF, showing that the test has high sensitivity. Of the remaining 4% that were not rejected, close inspection shows that three-quarters seem to contain true ORFs. Some define short ORFs with conserved start and stop codons in all four species, and others extend *S. cerevisiae* ORFs in the 5'- or 3'-direction in each of the other three species (see below). Thus, at most, 1% of true intergenic regions failed to be rejected by the RFC test.

Evaluating ORFs

We sought to apply the RFC test to all 6,062 ORFs in SGD. A total of 117 could not be analysed because they were almost completely contained within an overlapping ORF (99 cases, with average non-overlapping portion of 12 bp) or because an orthologous region could not be unambiguously defined in any of the species (18 cases).

Of the 5,945 ORFs tested, the analysis strongly validated 5,550 ORFs. The vote was unanimous in 5,458 cases ($\sim 98\%$). In the remaining cases, a valid gene appears to have degenerated in one of the four species. A total of 367 ORFs were strongly rejected. These rejections were unanimous in 63% of cases. In most of the remaining cases, *S. paradoxus* was too closely related to *S. cerevisiae*

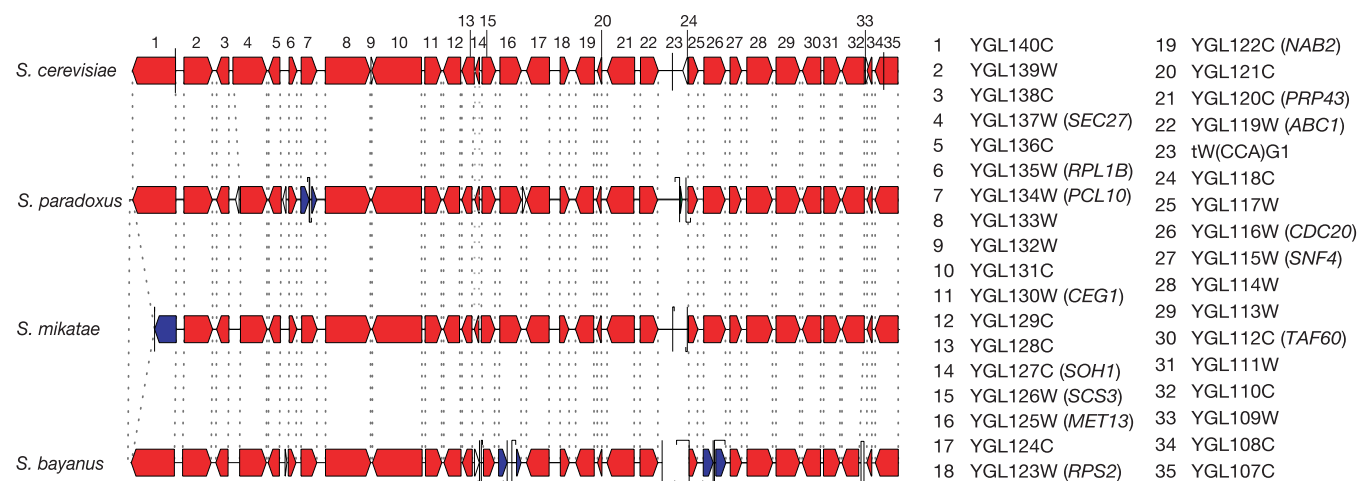


Figure 1 Aligned ORFs across four species. A 50-kb segment of *S. cerevisiae* chromosome VII aligned with orthologous contigs from each of the other three species. Predicted ORFs are shown as arrows pointing in the direction of transcription. Orthologous ORFs are connected by dotted lines and are coloured by the type of correspondence: red

for 1-to-1 matches, blue for 1-to-2 matches and white for unmatched ORFs. Sequence gaps are indicated by vertical lines at the ends of contigs, with the estimated size of each gap shown by the length of the hook. See Supplementary Information for 250 such figures tiling the complete *S. cerevisiae* genome.

to have accumulated enough frameshifts to allow definitive rejection.

Only 15 of these rejections involve one of the 3,966 named ORFs. We carefully inspected these 15 ORFs (*KRE20*, *KRE21*, *KRE23*, *KRE24*, *VPS61*, *VPS65*, *VPS69*, *BUD19*, *FYV1*, *FYV2*, *FYV12*, *API2*, *AUA1*, *ICS3*, *UTR5* and *YIM2*) and concluded that all were indeed likely to be spurious. Most lack experimental evidence. For the remainder, reported phenotypes associated with deletion of the ORF seem likely to be explained by the fact that the ORF overlaps the promoters of other known genes. The method thus rejects few known genes, indicating that it has high specificity.

Most of the rejections (352, or 96%) involve uncharacterized ORFs. Most of these (209) overlap another well-conserved ORF, but

show many insertions and deletions in the non-overlapping portion. The remainder (147) tend to be small (median of 111 amino acids, with 93% ≤ 150 amino acids) and show atypical codon usage^{1,29,30}. With only one exception, SGD reports no compelling biological evidence (such as changes in mRNA expression) to suggest that these ORFs encode a true gene. The one exception is YBR184W, which appears to represent a true gene that fails the RFC test because it is evolving very rapidly (see below).

The analysis deadlocked (one confirmation, one rejection, one abstention) for 28 ORFs (0.5%). Together with the 117 cases that could not be analysed, the RFC analysis thus produced clear results for all but 145 ORFs. We inspected these cases individually and found convincing evidence (on the basis of conservation of amino

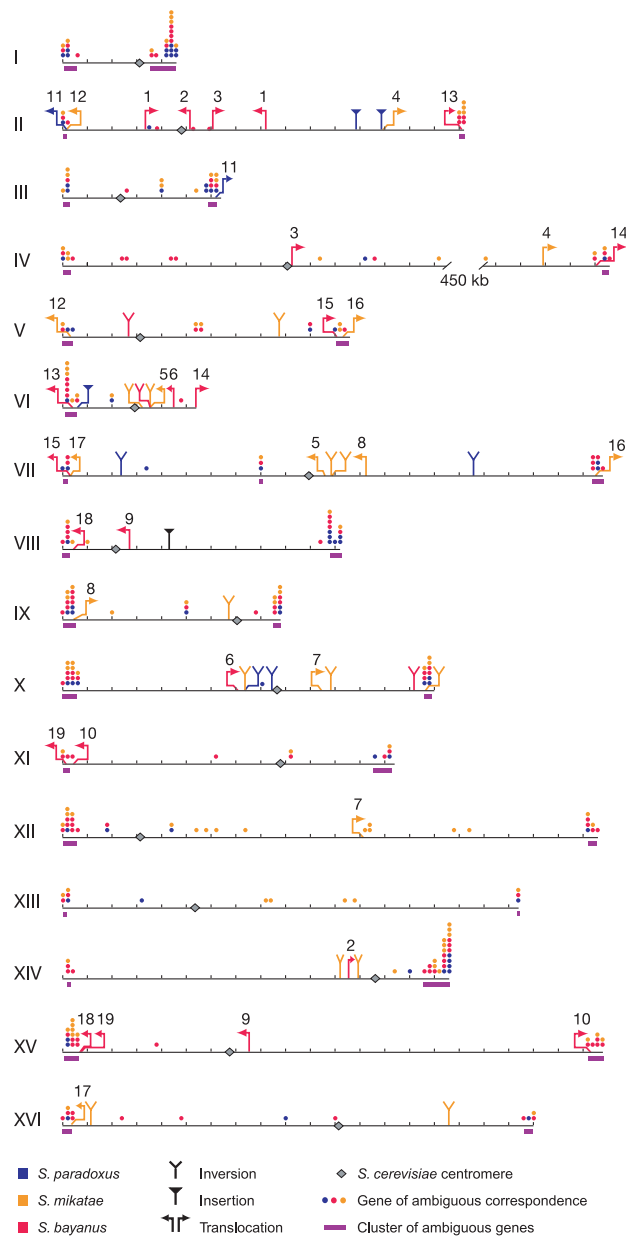


Figure 2 Genome evolution. Locations in the *S. cerevisiae* genome of chromosomal rearrangements in each of the three species are shown by vertical bars with the symbol at the top indicating the type of rearrangement (inversions, insertions and translocations) and colour indicating the species. Tick marks are shown every 50 kb. Inverted segments are small and their endpoints are indistinguishable at this scale. All inversions are flanked by tRNA genes in opposite transcriptional orientation. Reciprocal translocations involve exchange between two breakpoints; the matching pairs are indicated by their common

numbering and the orientation of the arrowheads. All non-telomeric translocations are mediated by Ty elements (2, 4, 5, 6, 7, 8, 9) or nearly identical copies of ribosomal protein genes (1, 3). Location in the *S. cerevisiae* genome of genes whose correspondence in the other species is ambiguous is shown by dots, with colour indicating the species in which the ambiguity occurs. A gene whose correspondence is ambiguous in all three species thus appears as three dots of different colours. Clusters of ambiguous genes are denoted by purple horizontal bars; 31 of 32 clusters occur at telomeres.

acids, start and stop codons, and the presence of indels) that 20 are valid protein-coding genes and 105 are spurious. We were unable to reach a judgment in the remaining 20 cases.

Finally, we performed certain consistency checks on the accepted ORFs. In 32 cases, two adjacent ORFs in *S. cerevisiae* are joined into a single ORF in all three other species. In every case, a single nucleotide change would suffice to join the ORFs in *S. cerevisiae* (either a substitution altering a stop codon or an indel altering the reading frame). In principle, these cases could represent errors in the genome sequence, mutations private to the sequenced strain S288C, or substitutions fixed in *S. cerevisiae*. We examined 19 cases by resequencing the relevant region in S288C. Our results revealed an error in the published sequence in 11 cases (establishing that there is a single ORF in S288C) and confirmed the published sequence in the remaining seven cases. Sequencing of additional strains will be required to determine whether these remaining cases represent differences in the S288C strain alone or in *S. cerevisiae* in general.

We also found two named ORFs (*FYV5* and *CWH36*) that pass the RFC test and cause phenotypes when deleted, but show no significant protein similarity across the four species. In both cases, inspection reveals that the opposite strand encodes a protein that shows strong amino-acid conservation. (The latter gene has two introns, increasing the count of doubly spliced genes to eight.) In each case, we postulate that the protein responsible for the reported deletion phenotype is encoded on the opposite strand.

Revised yeast gene catalogue

On the basis of the analysis above, we propose a revised yeast gene catalogue consisting of 5,538 ORFs encoding proteins of ≥ 100 amino acids. This reflects the proposed elimination of 503 ORFs (366 from the RFC test, 105 by manual inspection and 32 through merger). A total of 20 ORFs in SGD remain unresolved. Complete information about the gene catalogue is provided in Supplementary Information and will be discussed more fully in a subsequent manuscript in collaboration with SGD and other yeast investigators. The revised gene count is consistent with at least two recent predictions based on light shotgun coverage of related species^{5,6}.

We believe that this represents a reasonably accurate description of the yeast gene set, because the analysis examines all ORFs encoding ≥ 100 amino acids, the methodology has high sensitivity

and specificity, and the evidence is unambiguous for most of the ORFs. Nonetheless, some errors are likely to remain. The results could be confirmed and remaining uncertainties resolved by sequencing of additional related yeast species, as well as by other experimental methods.

Analysis of small ORFs

SGD also lists 141 ORFs encoding 50–99 amino acids for which some biological evidence has been published. Applying the RFC test and inspecting the results, we conclude that 120 seem to be true genes, 18 seem to be spurious ORFs and three remain unresolved.

We systematically searched the remainder of the *S. cerevisiae* genome and evaluated all ORFs in this size range. Control experiments demonstrated that the RFC test has high power to discriminate reliably between valid and spurious ORFs in this size range (see Supplementary Information). The genome contains 3,161 such ORFs; nearly all are readily rejected by the RFC test. However, 43 new genes were identified. These ORFs not only pass the RFC test, but they also have orthologous start and stop codons (see Methods). Five of these have been reported in the literature subsequent to the SGD release studied here.

SGD also lists 32 ORFs encoding < 50 amino acids. We did not undertake a systematic search for all such ORFs, because control experiments showed that the RFC test lacked sufficient power to prove the validity of such small ORFs (see Supplementary Information). However, it is able to reject seven of the 32 ORFs as probably spurious. Our yeast gene catalogue thus contains 188 short genes (encoding < 100 amino acids), of which 43 are new.

Defining gene structure

Comparative genome analysis not only improves the recognition of true ORFs, it also yields much more accurate definitions of gene structure—including translation start, translation stop and intron boundaries.

Previous annotation of *S. cerevisiae* has defined the start of translation as the first in-frame ATG codon. However, the actual start of translation could lie 3' to this point or (if sequencing errors or mutations have obscured an earlier in-frame ATG codon) 5' to this point. We identified 210 cases in which the presumed translational start in *S. cerevisiae* does not correspond to the first in-frame start codon in at least two of the three other species (see Methods). In most of these cases, inspection of the sequence alignments provides strong evidence for an alternative conserved position for the translational start, either 3' or 5' to the previous annotation (Fig. 5a, b). We examined four cases in which the comparative data suggested an earlier start codon and found, by resequencing, that all correspond to errors in the published sequence of S288C.

Similarly, we identified 330 cases in which the presumed translational stop codon in *S. cerevisiae* does not correspond to the first in-frame stop codon in at least two of the three species. In about 25% of these cases, the other three species share a common stop

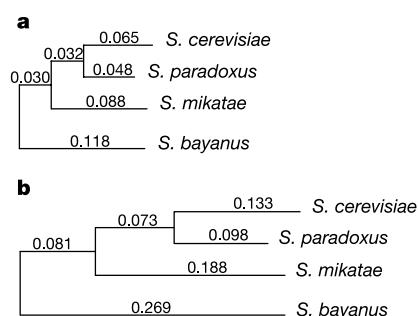


Figure 3 Evolutionary tree of the four yeast species. Branch lengths denote number of substitutions per site in genic (a) and intergenic (b) sequence. Pairwise distances were calculated using the Kimura two-parameter model and trees were constructed using the UPGMA method. The relative rate of the closest species was computed using the neighbour-joining method. On the basis of single-species changes in four-way alignments, we found *S. paradoxus* to evolve 67% slower than *S. cerevisiae* in coding regions and 70% slower in intergenic regions. In coding regions, the average nucleotide per cent identity to *S. cerevisiae* is 90% for *S. paradoxus*, 84% for *S. mikatae* and 80% for *S. bayanus*. In intergenic regions, the nucleotide per cent identity is, respectively, 80%, 70% and 62% for the three species. Additional information can be found in Supplementary Information. The nucleotide substitution level in aligned intergenic nucleotides between *S. cerevisiae* and *S. bayanus* corresponds roughly to that between human and mouse. Ribosomal DNA sequence analysis suggests divergence times of 5–20 million years for the four species.

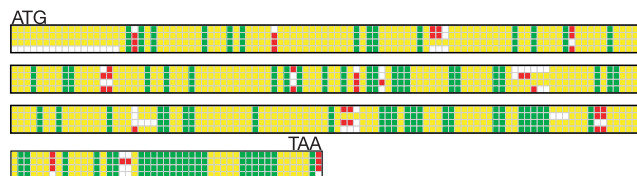


Figure 4 Spurious ORF rejected by RFC test. Schematic representation of the multiple sequence alignment of ORF YDR102C. Aligned nucleotides across the four species are shown as stacked squares (*S. cerevisiae*, *S. paradoxus*, *S. mikatae* and *S. bayanus*, respectively) coloured by their conservation: green for conserved positions, yellow otherwise. Alignment gaps are shown in white and frame-shifting insertions (length not a multiple of 3) are shown in red. In addition to the abundance of frame-shift indels shown here, numerous in-frame stop codons are observed in the other three species. Full nucleotide sequence alignment is provided in Supplementary Information.

codon, and a single base change to the *S. cerevisiae* sequence would result in a stop codon in the corresponding location (Fig. 5c, d). We examined 17 such cases and found that 15 are explained by errors in the published sequence of S288C. The remaining 75% of cases seem to represent true differences in the location of the translational stop across the species. Thus, stop codons seem to show more evolutionary variability in position than start codons.

New introns

We also examined the conservation of introns in the yeast genome. We studied 218 of the 240 ORFs reported in SGD to contain at least one intron (omitting the rest primarily due to lack of an orthologous alignment).

In 92% of cases, the donor, branchpoint and acceptor sites were all strongly conserved with respect to both location and sequence. Moreover, exon boundaries closely demarcated the domains of sequence conservation as measured by both nucleotide identity and absence of indels.

Discrepancies were found in 17 cases, of which at least nine strongly suggest that the previous annotation is incorrect. Five identify a new first exon (Fig. 5e) and four predict that a previously annotated intron is spurious.

We then sought to identify previously unrecognized introns by searching the *S. cerevisiae* genome for conserved splicing signals (see Methods) and then visually inspecting each case. We predict 58 new introns. Fifty cases affect the structure of known genes (defining new 5' exons in 42 cases, 3' exons in seven cases and an internal splice in one case) and two indicate the presence of new genes. The relationship of the apparent splice signals to existing genes is unclear for the remaining six cases.

We compared our predictions to the results of experimental studies by Ares and colleagues to identify new introns using techniques such as microarray hybridization³¹. Of our 58 predicted introns, 20 were independently discovered by this group. Of the four

annotated introns predicted to be spurious, all four show no experimental evidence of splicing. Our remaining predictions are currently being tested in collaboration with Ares and colleagues.

Despite the intensive study of *S. cerevisiae* so far, comparative genome analysis points to the need for a major revision of the yeast gene catalogue, affecting more than 15% of all ORFs. Moreover, the results suggest that comparative analysis of a modest collection of species can permit accurate definition of genes and their structure.

Rapid and slow evolution of genes

Comparative genomic approaches for gene identification require that orthologues can be recognized and aligned in most of the related species. Difficulties could arise in cases of rapid evolution, including the acquisition or loss of entire genes, the rapid divergence of nucleotide sequence or the presence of large insertions. We characterized such events in the four yeast species.

Species-specific genes

We noted above that *S. cerevisiae* contains 18 genes for which we could not identify orthologues in any of the other species, of which seven encode ≥ 200 amino acids. These may be species-specific genes in *S. cerevisiae*, but alternatively could simply reflect gaps in the available draft genome sequences.

This uncertainty does not arise, however, in the reverse direction in identifying genes in the related species that lack an orthologue in *S. cerevisiae*. We found a total of 35 such ORFs encoding ≥ 200 amino acids (with the minimum length chosen to ensure that these are likely to represent valid genes). The list includes five genes unique to *S. paradoxus*, eight genes unique to *S. mikatae* (two of which are 99% identical) and 19 genes unique to *S. bayanus* (three of which form a gene family with $\geq 90\%$ pairwise identity). There is also one gene represented by orthologous ORFs found in the latter

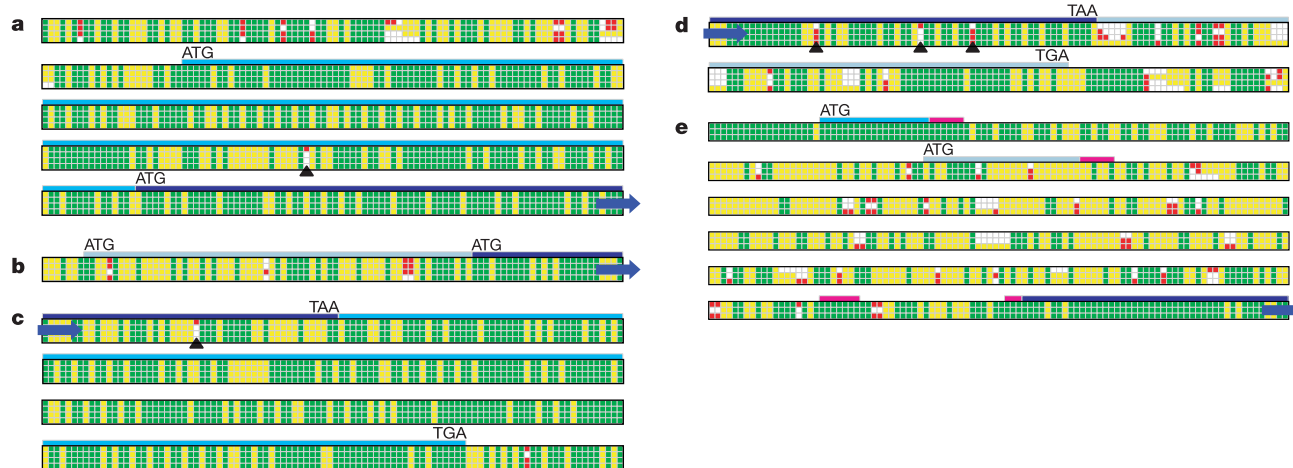


Figure 5 Examples of proposed changes in gene structure. Representation of alignments as in Fig. 4 (see Supplementary Information for full nucleotide alignments). Exon locations are indicated by overhead bars. Dark blue bars indicate portions of annotated exons unchanged in our analysis, and dark blue arrows indicate the continuation of ORFs beyond the alignment shown; light blue bars indicate new 5' or 3' extensions of exons (in **a** and **c**) and a new first exon (in **e**); grey bars indicate rejected portions of previously annotated exons (in **b** and **d**). Black triangles indicate the predicted errors in the published sequence of S288C; these predictions were confirmed by resequencing. In all five cases the proposed change leads to perfect agreement in ORF boundaries across all four species. **a**, Start moved upstream for *SAP155*. Elimination of a single base in the *S. cerevisiae* sequence extends the ORF 74 amino acids in the 5' direction, with a conserved start codon. Resequencing confirmed that the published sequence of S288C is in error. **b**, Start moved downstream for *YBP14*. No change to the *S. cerevisiae* sequence is required. The comparative data suggest that the true translation start is not until the second in-frame

ATG: the first ATG is not conserved and multiple frame-shifting indels are found in the rejected portion. **c**, Stop moved downstream for *YJL160C*. Elimination of a single base in the *S. cerevisiae* sequence extends the ORF by 108 amino acids in the 3' direction, with an orthologous stop codon in all four species. Resequencing confirmed that the published sequence of S288C is in error. **d**, Stop moved upstream for *YIP4*. Comparative data suggest a shorter ORF with an orthologous stop codon in all four species. The previously annotated stop codon is not conserved, and multiple frame-shifting indels are found in the rejected portion. All three predicted errors in the published sequence of S288C were confirmed by resequencing. **e**, Changed first exon for *RPL26B*. No change to the *S. cerevisiae* sequence is required. *S. paradoxus* sequence not shown. Conservation of splice donor and branchpoint suggests an alternative to the previously annotated exon 1 that shows no conservation of start codon or splice donor and that has accumulated numerous mutations and frame-shifting indels.

two species only, and one represented by orthologous ORFs in all three related species.

These species-specific ORFs are notable with respect to both function and location. Most (63%) can also be assigned biological function on the basis of strong protein-sequence similarity with genes in other organisms. Most involve sugar metabolism and gene regulation (including one encoding a silencer protein). The majority (69%) are found in telomeric regions and an additional set (17%) are immediately adjacent to Ty elements; these locations are consistent with rapid genome evolution.

A coincidence was noted in the region between YFL014W and YFL016W in *S. cerevisiae*. In the orthologous regions in all four species, we find a species-specific ORF in every case (165, 111, 136 and 228 amino acids), but these four ORFs show little similarity at the protein level. The amino-acid sequence has been disrupted by frame-shifting indels, but a long ORF has been maintained in each case. The explanation for this phenomenon is unclear, but may prove interesting.

Rapidly evolving genes

The gene analysis described above rejected only a single ORF (YBR184W) that is clearly biologically meaningful. The explanation seems to be that the gene is evolving rapidly. The region containing YBR184W corresponds to a large ORF in all four species (524, 558, 554 and 556 amino acids), but the alignment shows unusually low sequence conservation. The sequence has only 32% nucleotide identity and 13% amino-acid identity across the four species. Pairwise alignments across the species show numerous insertions and deletions, explaining why the gene failed the RFC test. (Notably, multiple alignment of all four species simultaneously improves the alignment sufficiently to allow the gene to pass the RFC test; this suggests a way to improve the test.)

The rapid divergence is suggestive of a gene under strong positive selection. We tested this hypothesis by calculating the K_a/K_s ratio (the normalized ratio of amino-acid-altering substitutions to silent substitutions), a traditional test for positive selection³². Whereas typical genes in *S. cerevisiae* show a K_a/K_s ratio of 0.11 ± 0.02 , YBR184W has a ratio of 0.689. This ratio ranks as the third highest observed among all yeast genes (if three small domains with high conservation are excluded, the ratio rises to 0.774). The two genes with higher K_a/K_s ratio are YAR068W, a putative membrane protein, and YER121W, whose expression changes under stress.

The protein encoded by YBR184W has not been studied extensively, but expression studies show that the gene is induced during sporulation³³, and sequence analysis shows that it is similar to the gene *YSW1*, which encodes a spore-specific protein. This is consistent with the observation that many of the best-studied examples of positive selection in other organisms are genes related to gamete function.

Evidence of rapid protein change

Most of the nucleotide changes in protein-coding regions are silent or affect individual amino acids. However, a small number of events suggest additional mechanisms of rapid protein change. These events include closely spaced compensatory indels that affect the translation of small contiguous amino-acid stretches. They also include the loss and gain of stop codons (by a nucleotide substitution or a frame-shifting indel), which may result in the rapid change of protein segments or the translation of previously non-coding regions³⁴. Such events are observed more frequently near telomeric regions and may affect silenced genes or recently inactivated pseudogenes.

Furthermore, we found a small number of differences in the length of orthologous proteins. These typically involve changes in the copy number of tri-nucleotide repeats, such as (CAA), which encodes hydrophobic stretches often involved in protein-protein interactions. The most marked example is seen for the *TFPI* gene, which encodes a vacuolar ATPase. The *S. cerevisiae* gene contains an

insertion of 1,400 bp that is absent in the three related species. The insertion corresponds to the recent horizontal transfer of a known post-translationally self-splicing intein, *Vma1* (ref. 35).

Slowly evolving genes

To complement the analysis of rapidly changing genes, we scanned the genome for genes with an unusually slow rate of evolution. One case stands out as an extreme outlier: the mating-type gene *MATa2*. The gene shows perfect 100% conservation at the amino-acid level over its entire length (119 amino acids) across all four species. More notably, the gene shows perfect 100% conservation at the nucleotide level as well (357 bp). This differs sharply for the typical pattern seen for protein-coding genes, which show relaxed constraint in third positions of codons.

Notably, the *MATa2* gene is the only one of the four mating-type genes (the others being *MATa1*, *MATa2* and *MATa1*) whose biochemical function remains unknown despite two decades of research³⁶. An important clue may be that the sequence of *MATa2* is identical in all four species to the 3' end of the *MATa2* gene. Perfect conservation at the nucleotide level and identity to the terminus of *MATa2* suggests that *MATa2* may function not by encoding a protein, but rather by encoding an antisense RNA or a DNA site.

Variation in the rate of genome evolution seems not to pose a serious problem for gene identification for the species studied here. Various examples of extremely rapid or slow gene evolution can be identified, however, and each is likely to merit further study.

Genome-wide identification of regulatory elements

Direct identification of regulatory elements is more challenging than for genes. Such elements are typically short (6–15 bp), tolerate some degree of sequence variation and follow few known rules. So far, most have been found by experimental manipulation, such as systematic mutation of individual promoter regions; however, the process is laborious and unsuited for genome-scale analysis.

Computational analysis of single genomes has been successfully used to identify regulatory elements associated with known sets of related genes^{8–10}, but these approaches do not have sufficient power to permit comprehensive direct identification of regulatory elements³⁷.

Comparative genomics offers various approaches for finding regulatory elements. The simplest is to perform cross-species sequence alignment to find 'phylogenetic footprints'—regions of unusually high conservation. This approach has long been used to study promoters of specific genes in many organisms^{11,13,24,38,39}, and recently was applied across the entire human and mouse genomes⁷. The genome alignments of the four *Saccharomyces* species can similarly be used to study each yeast gene, to help define promoters and other islands of intergenic conservation (Fig. 6).

Our interest was to go beyond inspection of individual islands of conservation to construct a comprehensive dictionary of regulatory elements used throughout the genome. We investigated the conservation properties of known regulatory motifs and used the insights gained to design an approach for *de novo* discovery of regulatory motifs directly from the genome. We then developed an approach for inferring a candidate function of these motifs, making use of biological knowledge about genes.

Conservation of the Gal4-binding site

We first studied the binding site for one of the best-studied transcription factors, Gal4, whose sequence motif is CGGn₍₁₁₎CCG (which contains 11 unspecified bases). Gal4 regulates genes involved in galactose metabolism, including the *GAL1* and *GAL10* genes, which are divergently transcribed from a common intergenic region (Fig. 6). The Gal4 motif occurs three times in this intergenic region,

and all three instances show perfect conservation across the four species. In addition, there is a fourth experimentally validated binding site⁴⁰ for Gal4 that differs from the consensus by one nucleotide in *S. cerevisiae*. This variant site is also perfectly preserved across the species.

We then examined the frequency and conservation of Gal4-binding sites across the aligned genomes. In *S. cerevisiae*, the Gal4 motif occurs 96 times in intergenic regions and 415 times in genic (protein-coding) regions. The motif displays certain marked conservation properties (see Methods). First, occurrences of the Gal4 motif in intergenic regions have a conservation rate (proportion conserved across all four species) that is about fivefold higher than for equivalent random motifs (12.5% compared with 2.4%). Second, intergenic occurrences of the Gal4 motif are more frequently conserved than genic occurrences (12.5% compared with 3%). By contrast, random motifs are less frequently conserved in intergenic regions than in genic regions (3.1% compared with 7.0%), reflecting the lower overall level of conservation in intergenic regions. Thus, the relative conservation rate in intergenic compared with genic regions is about 11-fold higher for Gal4 than for random motifs. Third, the Gal4 motif shows a higher conservation rate in divergent compared with convergent intergenic regions (those that lie upstream compared with downstream of both flanking genes); no such preferences are seen for control motifs. These three observations suggest various ways to discover motifs based on their conservation properties (see conservation criteria below).

Catalogue of known motifs

We extended these observations by assembling a catalogue of 55 known regulatory sequence motifs (Table 2), by starting with two public databases (SCPD⁴¹ and YTFD⁴²) and curating the entries to select those with the best support in the literature. Nearly all of these sequence motifs are binding sites of known transcription factors.

We defined a motif conservation score (MCS) on the basis of the conservation rate of the motif in intergenic regions: the MCS is measured in standard deviations above the rate for comparable control motifs (see Methods). Most of the known motifs show strong conservation, with 60% having MCS ≥ 4 (which is substantially higher than expected by chance). Some of the motifs, however, show relatively modest MCS. These motifs may be incorrect, suboptimal or not well conserved.

Methodology for genome-wide motif discovery

Our approach involves first identifying conserved 'mini-motifs' and then using them to construct full motifs (see Methods). Mini-motifs are sequences of the form XYZ_{n(0-21)}UVW, consisting of two triplets of specified bases interrupted by a fixed number (from 0 to 21) of unspecified bases. Examples are TAGGAT, ATAnnGGC, or the Gal4 motif itself. The total number of distinct mini-motifs is 45,760, if reverse complements are grouped together.

Conserved mini-motifs are then defined according to three conservation criteria (CC1–3), based on our observations about

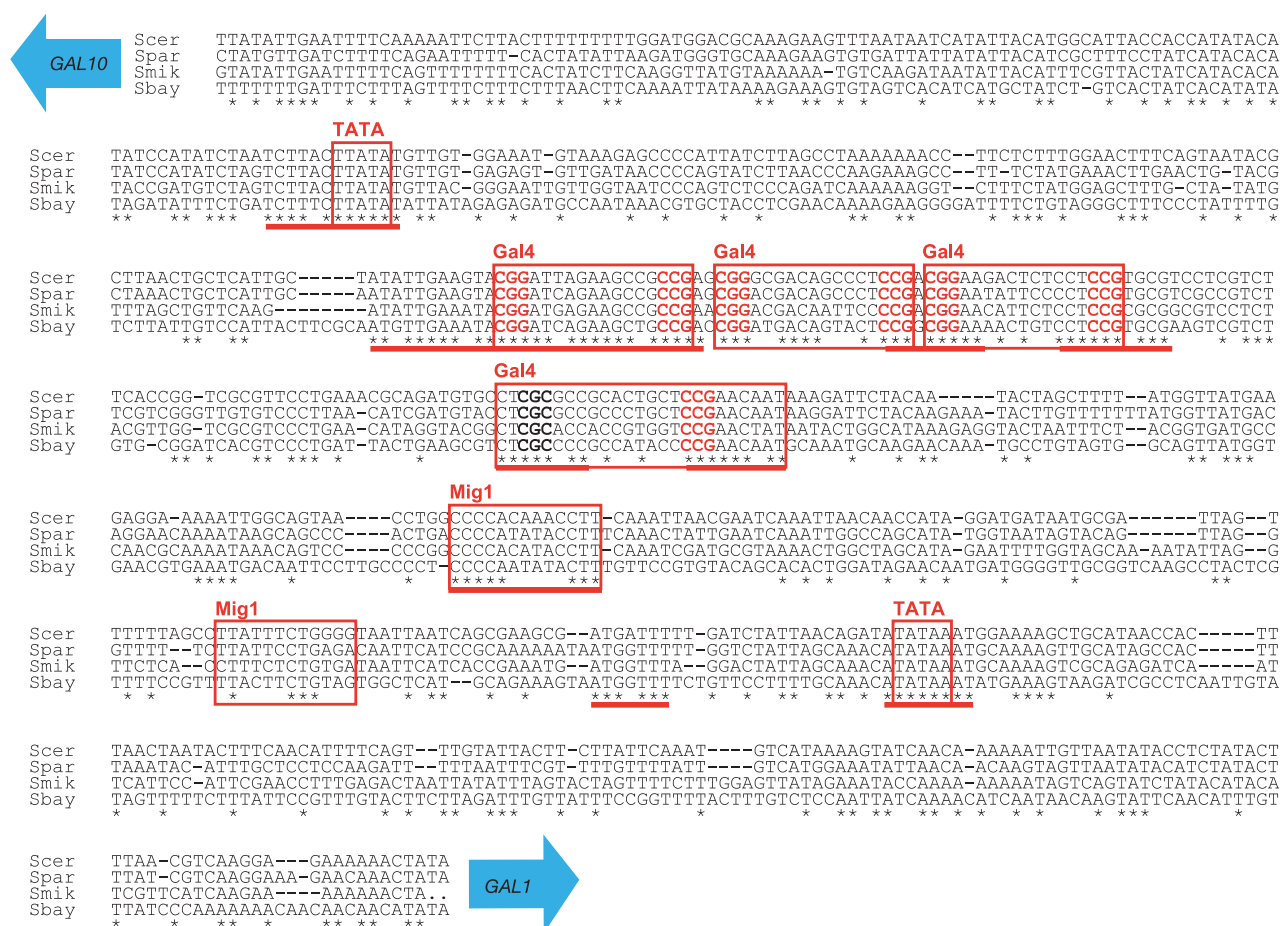


Figure 6 Conservation in the *GAL1-GAL10* intergenic region. Multiple alignment of the four species shows a strong overlap between functional nucleotides and stretches of conservation. Asterisks denote conserved positions in the multiple alignment. Blue arrows denote the start and transcriptional orientation of the flanking ORFs. Experimentally validated factor-binding footprints are boxed and labelled according to the bound factor.

Stretches of conserved nucleotides are underlined. Nucleotides matching the published Gal4 motif are shown in red. The fourth experimentally validated site differs: it shows a longer footprint and a non-standard consensus motif (bold). This variant motif is also conserved across all four species. Scer, *S. cerevisiae*; Spar, *S. paradoxus*; Smik, *S. mikatae*; Sbay, *S. bayanus*.

the properties of the Gal4 motif. In each case, conservation rates are normalized to appropriate random controls. The conservation criteria are: (1) intergenic conservation (CC1), the mini-motif shows a significantly high conservation rate in intergenic regions; (2) intergenic-genic conservation (CC2), the mini-motif shows significantly higher conservation in intergenic regions than in genic regions; (3) upstream-downstream conservation (CC3), the mini-motif shows significantly different conservation rates when it occurs upstream compared with downstream of a gene.

The conserved mini-motifs are then used to construct full motifs (Fig. 7). They are first extended by searching for nearby sequence positions showing significant correlation with a mini-motif. The extended motifs are then clustered, merging those with substantially

overlapping sequences and those that tend to occur in the same intergenic regions. Finally, a full motif is created by deriving a consensus sequence (which may be degenerate).

Each full motif is assessed for genome-wide conservation by calculating its MCS, and those motifs with MCS ≥ 4 are retained. Each full motif was also tested for enrichment in upstream compared with downstream regions, by comparing its conservation rate in divergent versus convergent intergenic regions.

Results of genome-wide motif discovery

Most of the 45,760 possible mini-motifs show no distinctive conservation pattern. However, approximately 2,400 mini-motifs show high scores by one or more of these criteria (Fig. 7a–c). There

Table 2 Known motifs and related discovered motifs

Factor	Known motif		Discovered motif*			
	Motif	MCS†	Motif	Genome-wide‡	Category-based§	MCS
ABF1	RTCRYnnnnnACG	50.0	RTCRYknnnnACGR	S	S	36.2
UME6	TCGGCGGCTA	20.9	TSGGCGGCTAWW	S	NC	23.4
CBF1	RTCACRTG	19.0	RTCACGTGV	S	S	17.6
NDT80	TCGGCGGCTDW	18.6	TSGGCGGCTAWW	S	NC	23.4
REB1	TTACCCGG	17.8	RTTACCCGRM	S	S	34.3
MCM1a	TTWCCnWWWRGGAAA	16.5	TTCCnaAttnGGAAA	S	S	13.8
SWI6	ACGCGT	16.4	WCGCGTCGCGt	S	S	10.2
PHO4	CACGTG	16.1	RTCACGTGV	S	S	17.6
MBP1	ACGCGTnA	14.8	WCGCGTCGCGt	S	S	10.2
SWI4	TTTTCGCG	12.4	WTTTCGCGTT	S	S	12.0
DAL81	GATAAG	12.1	–	–	NE	–
RPN4	TTTTGCCACC	11.5	TTTTGCCACCG	S	NC	11.0
MSN2	CCCT	11.3	hRCCCYTWDt	S	NE	7.8
MSN4	CCCT	11.3	hRCCCYTWDt	S	NE	7.8
PDR1	CCGCGG	9.3	YCCGSGGS	S	NE	6.7
ESR2	AAAATTTT	8.9	GRRAAAWTTTCACT	S	NC	15.6
MIG1	CCCCRSWWWW	8.7	DCCCCGCGH	S	NE	8.2
MIG1b	CCCGCG	8.4	DCCCCGCGH	S	NE	8.2
BAS1	TGACTC	8.3	ATGACTCWT	S	S	6.1
GCN4	ATGACTCAT	8.2	ATGACTCWT	S	S	6.1
GAL4	CGGnnnnnnnnnnCCG	8.0	CGGnnnMGnnnnnnnnCGC	S	S	5.0
HSF1b	TTCTAGAA	7.8	TTCTMGAAAG	S	S	7.0
ESR1	GATGAG	7.7	gcGATGAGmtgaraw	S	NC	24.7
MET31	AAACTGTGGC	6.8	SKGTGGSgc	S	S	8.1
AFT1	YRCACCCR	6.8	RVACCCTD	S	NC	10.3
TEA1	CGGnCGG	6.8	–	–	NC	–
PUT3	CGGnnnnnnnnnnCCG	6.2	CCGMnnnnnnnnnnmSGR	W	NE	5.4
HAP2	TGATTGGC	5.7	TGATTGGT	–	S	[6.4]
RAP1	ACACCCATACATTT	5.2	ACACCCACACATnnC	S	S	9.9
LEU3	CCGGnnCCGG	4.9	CCSGTAnCCG	S	S	6.5
MCM1b	YTTCCTAATTWGnnCn	4.8	TTCCnaAttnGGAAA	S	S	13.8
INO4	CATGTGAAAT	4.1	GnnnCATTGTGAA	–	S	[6.8]
INO2	CATGTGAAAT	4.1	CATGTG	–	S	[4.4]
GLN3	GATAAK	3.8	–	–	NE	–
ADR1	GGAGA	3.7	–	–	NE	–
FKH2	TTGTTTACST	3.6	tTTGTTTACnTTT	S	S	10.8
FKH1	TTGTTTACST	3.6	tTTGTTTACnTTT	S	S	10.8
RLM1	CTAWWWWTAG	3.6	CTAnnTTTAG	S	S	[4.7]
SWI5	KGCTGR	3.4	TGCTGG	–	S	[6.1]
HAP1	CGGnnnTAnCCG	2.5	GCnnTTAnCCG	S	NC	4.8
XBP1	MCTCGARRRnR	2.5	TCTCGARRA	S	NC	12.5
MAC1	TTTGCTCA	2.3	TGCTCA	–	S	[5.4]
TBF1	TTAGGG	2.3	GKBAGGGT	S	NC	4.8
MSE	TTTTGTG	1.4	TTTTGTGTCRC	S	NC	9.9
STE12	RTGAAACA	0.7	YTGAAACA	–	S	[12.2]
DIG1	RTGAAACA	0.7	YTGAAACA	–	S	[12.2]
MET4	TGGCAATG	0.7	CGGTGGCAAAA	S	NE	–
HAP4	TnRTTGGT	0.5	TGATTGGT	–	S	[6.4]
SMP1	ACTACTAWWWWTAG	0.4	–	–	NE	–
ACE2	GCTGGT	–0.6	TGCTGGT	–	S	[7.4]
YAP1	TTACTAA	–1.1	–	–	NE	–
CIN5	TTACTAA	–1.1	–	–	NE	–
RME1	GAACCTCAA	–1.4	–	–	NE	–
HAC1	CAGCGTG	–1.4	–	–	NC	–
GCR1	GGAAG	–18.5	GGAAGC	–	S	[4.4]

Lower-case characters denote lower stringency. Degenerate nucleotides as follows: S = CG, W = AT, R = AG, Y = CT, K = GT, M = AC, B = CGT, D = AGT, H = ACT, V = ACG, N = ACGT. Characters in bold correspond to matches between known motif and discovered motif.

*Discovered motif is the best genome-wide discovered motif, if one was found. Otherwise, it is the category-based motif found using the gene category corresponding to chromatin immunoprecipitation with the corresponding transcription factor, if this category was available.

†Negative MCS values indicate that the motif showed a weaker conservation than the random controls. Known motifs with low MCS values (<1.0) were only found in the category-based search.

‡S indicates that a strong genome-wide match was found to the known motif; W, a weak match to the known motif was found.

§S indicates that a strong category-based motif was found using the gene category for ChIP with the corresponding transcription factor; NE indicates that the known motif was not enriched in this gene category; NC indicates that the gene category was available. Note that a strong category-based motif was found for every known motif for which the appropriate gene category was available and the known motif was actually enriched in the category.

||MCS value of the genome-wide motif, if one was found. Category-based score shown in brackets otherwise.

is substantial overlap among the mini-motifs produced by the three criteria, with about 50% of those found by one criterion also found by another.

The conserved mini-motifs give rise to a list of 72 full motifs having $MCS \geq 4$ (Table 3). Most of the motifs show preferential enrichment upstream of genes, but six are enriched downstream of genes. These 72 discovered motifs, found with no previous biological knowledge, show strong overlap with 28 of the 33 known motifs having $MCS \geq 4$. They include 27 strong matches and one weaker match.

The 72 discovered motifs also contain matches to eight of the 22 known motifs with $MCS < 4$. In these cases, the comparative analysis identified closely related motifs that have higher conservation scores than the known motifs and occur largely at the same genes; these may represent a better description of the true regulatory element.

Comparative genomic analysis thus automatically discovered 36 motifs with matches to most of the known motifs (65% of the full set, 85% of those with high conservation). It also identified 42 additional ‘new’ motifs not found in our list of known motifs.

Inferring function of genome-wide motifs

We next developed ways to assign candidate functions to these discovered motifs by the genes adjacent to conserved occurrences of

the motif with known gene categories. For inspiration, we again used the Gal4 motif. Given the biological role of Gal4, we considered the set of genes annotated to be involved in carbohydrate metabolism (126 genes according to the Gene Ontology (GO)⁴³ classification) with the set of genes that have a Gal4-binding motif upstream. The intergenic regions adjacent to carbohydrate metabolism genes comprise only 2% of all intergenic regions, but 7% of the occurrences of the Gal4 motif in *S. cerevisiae* (3.5-fold enrichment) and 29% of the conserved occurrences across the four species (15-fold enrichment).

These results suggest that a function of the Gal4 motif could be inferred from the function of the genes adjacent to its conserved occurrences. Such putative functional assignments can be useful in directing experimentation for understanding the precise function of a motif. We therefore assembled a collection of 318 yeast gene categories based on functional and experimental data. These categories consist of 120 sets of genes defined with a common GO classification in SGD⁴³; 106 sets of genes whose upstream region was identified as binding a given transcription factor in genome-wide chromatin immunoprecipitation (ChIP) experiments⁴⁴; and 92 sets of genes showing coordinate regulation in RNA expression studies⁴⁵. To measure how strongly the conserved occurrences correlated with the regions upstream (or downstream) of a

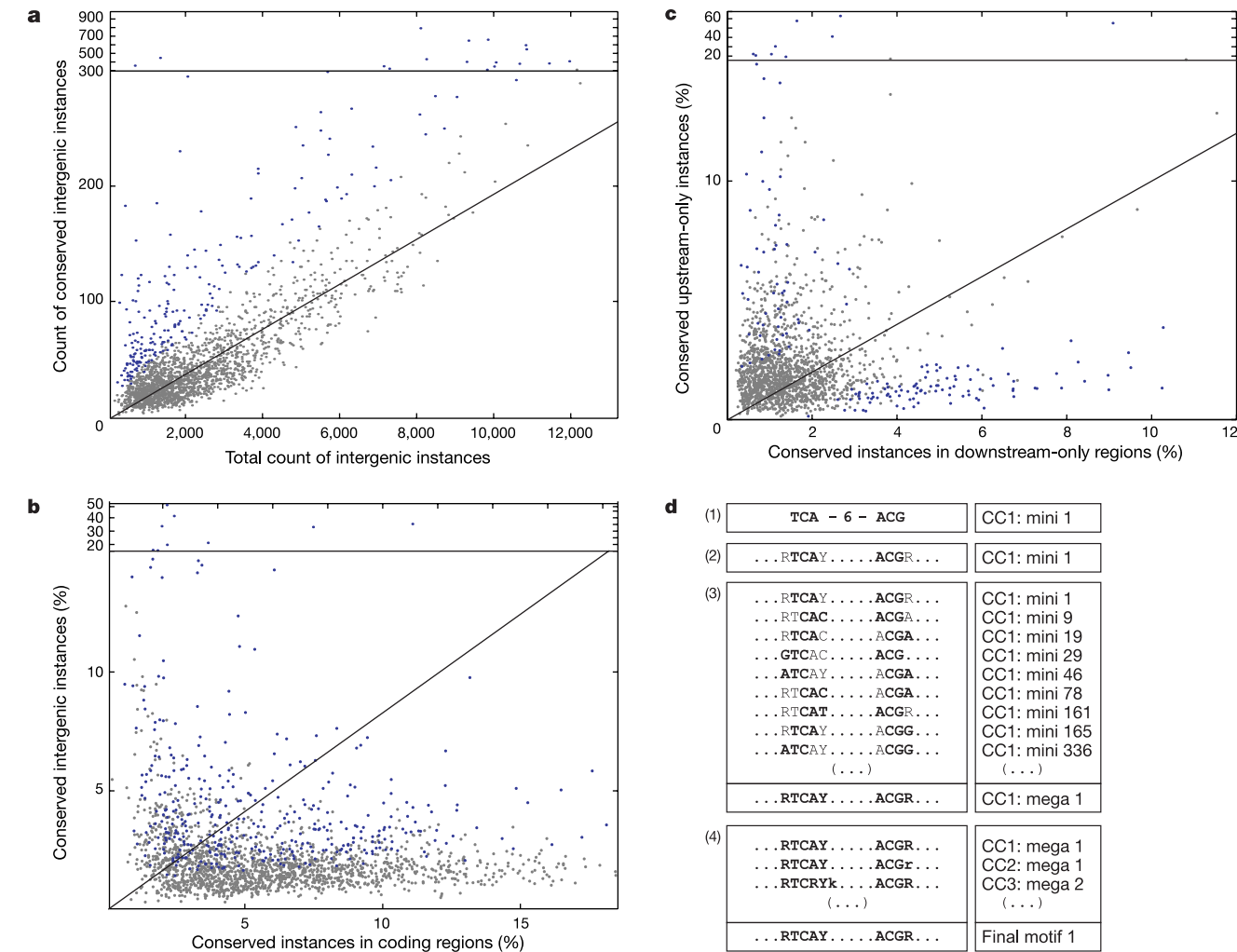


Figure 7 Distribution of motifs by conservation score. **a–c**, Points denote individual mini-motifs analysed with respect to the three conservation properties described in the text (CC1, intergenic conservation (**a**); CC2, intergenic–genic conservation (**b**); CC3, upstream–downstream conservation (**c**)). Blue points exceed the significance threshold used; grey points are non-significant. **d**, The generation of a full motif from mini-motifs:

(1) we selected mini-motifs using each conservation criterion; (2) we extended mini-motifs with additional conserved bases; (3) we generated mega-motifs by merging mini-motifs with similar extension; (4) we merged mega-motifs that frequently co-occur in the same intergenic regions.

particular gene category, we defined a category correlation score (CCS) (see Methods). A CCS ≥ 5 (nominal P -value $<10^{-5}$) is significant after accounting for testing of multiple categories.

Most of the 36 discovered motifs that correspond to known

motifs showed strong category correlation. Categories with the strongest correlation included those identified by ChIP with the transcription factor known to bind the motif, although many other relevant categories were identified. Of the 42 new motifs, 25 show

Table 3 **Discovered motifs**

No.	Discovered motif	Location*	MCS†	Best category‡	CCS§	Interpretation
1	YCGTnnnnmRYGAY	5'	36.2	ChIP: Abf1	90	Known: Abf1
2	RTTACCCGRM	5'	34.3	ChIP: Reb1	38	Known: Reb1
3	gcGATGAGmtgaraw	5'	24.7	Exp.: cluster 74	62	Known: Esr1 GATGAG
4	TSGGCGGCTAWW	5'	23.4	GO: meiosis	10	Known: Ume6/Ndt80
5	RTCACGTGV	5'	17.6	ChIP: Cbf1	27	Known: Cbf1/Pho4
6	WTATWTACADG	3'	17.4	Exp.: cluster 16 downstream	25	New: mitochondrial downstream
7	GRRAAAWTTTCACT	5'	15.6	Exp.: cluster 74	37	Known: Esr2
8	TTCCnaAttnGGAAA	5'	13.8	ChIP: Mcm1	29	Known: Mcm1
9	CGTTTCTTTTTCY.	5'	13.5	GO: filamentation	7	New: filamentation
10	TYTTCGAGA.	5'	12.5	Exp.: cluster 86	5	Known: Xbp1 (Hsf1-co-ocuring)
11	TTTTTCGCG	5'	12.0	ChIP: Swi4	21	Known: Swi4 fixed gap
11a	TTTT = CGCG¶	5'	12.0	ChIP: Swi4	—	New: Swi4 variable gap
12	TKACGCGTT	5'	12.0	ChIP: Mbp1	18	Known: Mbp1/Swi6
13	STGCGGnnnttTCTnnG	5'	11.8	GO: filamentation	11	New: filamentation
14	YCTATTGTT	5'	11.5	ChIP: Fkh2	6	New: Rlm1-like
15	TTTTGCCACCG	5'	11.0	GO: proteolysis	25	Known: Rpn4/Met4
16	tTTGTTTACnTTT	5'	10.8	ChIP: Fkh2	28	Known: Fkh1/2
17	RVACCTD	5'	10.3	—	—	Known: Aft1
18	WCGCGTCGCGt	5'	10.2	ChIP: Mbp1	17	New: double Mbp1
19	GGGTnACCC	5'	10.0	ChIP: Reb1	8	New: Reb1 palindrome
20	GnnATGTGTGGGTGT	5'	9.9	ChIP: Fhl1	5	Known: Rap1
21	TTTTGTGTCRC	5'	9.9	ChIP: Sum1	14	Known: Mse
22	TTTCAnCGCGC	5'	9.8	—	—	New: no category
23	TATTAWTATTATiMtnatta	3'	9.5	—	—	New: no category
24	SCGnHGGs	5'	8.8	GO: filamentation	6	New: filamentation
25	ACAGCCGCRY	5'	8.6	Exp.: cluster 37	6	New: expression cluster 37
26	DCGCGGGH	5'	8.1	Exp.: cluster 46	8	Known: Mig1b
27	SKGTGGSGc	5'	8.1	ChIP: Met31	5	Known: Met31
28	TTTTn(19)GCKCG	5'	7.8	—	—	Known: no category
29	HRCCCYTWDt	5'	7.8	Exp.: cluster 8	22	Known: Msn2/4
30	TKCCnnnnGGG	5'	7.3	ChIP: Mcm1	15	Known: Mcm1 (hits tRNA)
31	GTGTCAGTAAt	5'	7.1	ChIP: Sum1	15	New: Sum1
32	RGTTTTTCCG	5'	7.1	ChIP: Rgt1	7	New: Rgt1
33	TTCTMGAAGA	5'	7.0	ChIP: Hsf1	10	Known: Hsf1
34	YCCSGSGS	5'	6.7	GO: filamentation	9	New: filamentation
35	CnCTTTTATAC	5'	6.5	—	—	New: no category
36	CCSGTAnCGG	5'	6.5	ChIP: Leu3	8	Known: Leu3
37	SKTKCCTT	5'	6.4	GO: filamentation	7	New: filamentation
38	CTCCCCCTTAT	5'	6.4	Exp.: cluster 8	11	Known: Msn2/4
39	GCCCCG	5'	6.3	GO: filamentation	10	New: filamentation
40	SGCGCGRB	5'	6.3	—	—	New: no category
41	CTCSGCS	5'	6.2	—	—	New: no category
42	TGnKAGCGCCG	5'	6.2	—	—	—
43	ATGACTCWT	5'	6.1	ChIP: Gcn4	44	Known: Gcn4/Bas1
44	CCGAnnnTCGG	5'	6.1	Exp.: cluster 46	6	New: facilitators palindrome
45	SCGMnnnnnnKCG	5'	6.0	—	—	New: no category
46	CnCCGCGCnnTTTs	5'	6.0	—	—	New: no category
47	TTTTnnnnnnnnnnngGGGT	5'	5.8	—	—	New: no category
48	TGTRnCAW	3'	5.5	—	—	New: no category
49	YCSknnnnnnnnnKCGG	5'	5.4	Exp: cluster 46	6	Known: Put3
50	CGGnnnnnnnnnnnnKCGV	5'	5.4	—	—	New: no category
51	WGTGACg	5'	5.3	ChIP: Sum1	14	New: Sum1
52	RTCCCTV	5'	5.3	—	—	New: no category
53	YTCGTTTAGG	5'	5.2	GO: lipid metabolism	5	New: lipid metabolism
54	TYCGKRM	5'	5.2	GO: filamentation	7	New: filamentation
55	CGCnnnnnnnnnnnnBCGB	5'	5.1	—	—	New: no category
56	TWCCCCM	5'	5.0	Exp.: cluster 46	7	Known: Mig1 + facilitators
57	CGGCnnMGnnnnnnCGC	5'	5.0	ChIP: Gal4	7	Known: Gal4
58	CCGSnnnnnnGVC	5'	5.0	—	—	New: no category
59	TRTAMATAKWT	3'	4.8	ChIP: Dig1	7	New: Ste12 (hits tRNA)
60	TtATAnTATATAnA	3'	4.8	Exp: cluster 74 downstream	6	New: downstream cluster 74
61	GKBAGGGT	5'	4.8	GO: glycolysis	6	Known: Tbf1/new: glycolysis
62	GCnnTTAnCGG	5'	4.8	—	—	Known: Hap1
63	GGCSnnnnnnGnnnCGCG	5'	4.7	ChIP: Mbp1	6	Known: Mbp1-like
64	TTCTCnnnnnnCGC	5'	4.7	GO: filamentation	6	New: filamentation
65	SCGKnnnnnKCGD	5'	4.5	—	—	New: no category
66	AATATTCTT	3'	4.4	Exp.: cluster 46 downstream	5	New: downstream facilitators
67	CGCGTnnnnnnnnACG	5'	4.4	ChIP: Swi4	8	New: Swi4-vary gap
68	CCGHVGGM	5'	4.3	—	—	New: no category
69	CGCG = TTTT	5'	4.3	—	—	New: no category
70	CGCGnnnnnGGGS	5'	4.2	Exp.: cluster 46	6	New: expression cluster 46
71	CTGCAGGGR	5'	4.2	GO: filamentation	6	New: filamentation

Lower-case characters and degenerate nucleotides as in Table 2.

*Location indicates whether the motif tends to occur 5' or 3' of genes, based on preferential enrichment in divergent or convergent intergenic regions.

†Motif conservation score for discovered motif.

‡The gene category having the highest category correlation score (CCS) with discovered motif. Category types are: GO, Gene Ontology; ChIP, chromatin IP experiment; Exp., expression cluster. §CCS for best category.

||Interpretation of discovered motif, based on similarity to known motifs (Table 2) and category correlation. No category denotes that no category was enriched in the motif.

¶Motif 11a is closely related to motif 11 but shows a variable gap.

strong correlation with at least one category and thus can be assigned a suggestive biological function (Table 3).

Examples of new motifs

We describe a few of the new motifs below.

- Some motifs appear to define previously unknown binding sites associated with known transcription factors. Motif 32 is probably the binding site for Rgt1, which regulates genes involved in glucose transport⁴⁶; the motif occurs upstream of many such genes, including appearing five times upstream of *HXT1*, which encodes a high-affinity glucose transporter. Motifs 21, 31 and 51 are all associated with genes whose upstream regions are bound by Sum1, a transcriptional repressor of genes involved in meiosis. The first motif has been reported previously (Mse)⁴⁷, but the latter two are new and occur near genes whose products are involved in chromatin silencing and transcriptional repression.
- Some motifs do not match regions bound by known transcription factors, but show strong correlation with functional categories. Motif 9 occurs upstream of genes involved in nitrogen metabolism, including amino-acid and urea metabolism, nitrogen transport, glutamine metabolism and carbamoyl phosphate synthesis. Motif 25 is enriched among co-expressed genes (expression cluster 37) whose products function in vesicular transport and secretion, including GDP/GTP exchange factors essential for the secretory machinery, clathrin assembly factors and many vesicle and plasma membrane proteins. Motifs 9, 13, 24, 34, 37 may have a role in filamentation. They are all enriched in genes co-regulated during environmental changes, involved in signalling and budding, and bound by transcription factors involved in filamentation, such as Phd1.
- Six motifs show higher conservation downstream of ORFs. Some of these may be transcribed in the 3' untranslated region and have a regulatory role in mRNA localization or stability. The strongest (motif 6; see ref. 48) is found at genes whose product localizes to the cytosolic translational machinery, the mitochondrial DNA translational machinery or the mitochondrial outer membrane. Downstream motifs are also found enriched in a group of genes repressed during environmental stress (motif 60 with expression cluster 74) and a group of genes involved in energy production (motif 66 with expression cluster 46).
- Two motifs (motif 11a and motif 69) show variable gap spacing, suggesting a new type of degeneracy within the recognition site for a transcription factor complex. Motif 11a corresponds closely to the known motif for Swi4 (motif 11) but is interrupted by a central gap of 5, 7 or 9 bases; these variant motifs all show strong correlation with genes bound by Swi4 in ChIP experiments.

Category-based identification of regulatory elements

The motifs above were identified solely on the basis of overall conservation across the genome. We next explored whether additional motifs could be found by searching specifically for conservation within individual gene categories. We used the same procedure for constructing full motifs as above, but used a modified conservation criterion (CC4) that defines conserved mini-motifs as

those enriched in the intergenic regions of genes in the category (see Methods).

Known motifs

We first considered the 43 known motifs for which ChIP experiments had been performed with the transcription factor that binds the motifs⁴⁴. For each category defined by the ChIP experiment, we undertook category-based motif discovery.

Strong category-based motifs were found in 29 cases and these invariably corresponded closely to the known motifs (Table 2). These include 11 cases in which the motif had not been found by genome-wide motif discovery, suggesting that a category-based approach can be more sensitive in some cases.

No strong category-based motifs were found for the remaining 14 known cases, including seven cases in which genome-wide analysis yielded the known motif. Analysis of these 14 known motifs showed that none were, in fact, enriched in the ChIP-based category. This may reflect errors in the known motifs in some cases, and imperfect ChIP data in others. Genome-wide analysis may simply be more powerful than category-based analysis in some instances.

In all, 46 of the 55 known motifs were found by either genome-wide or category-based analysis. The remaining nine cases may reflect true failures of the comparative genomic analysis or errors in the known motifs.

New motifs

We then applied the approach to all of the gene categories. A total of 181 well-conserved motifs were identified, with many of these being equivalent motifs arising from multiple categories. Merging such motifs resulted in 52 distinct motifs, of which 43 were already found by the analyses described above. The remaining nine motifs represent new category-based motifs (Table 4), including the following.

- Three new motifs are associated with genes that are bound by the transcription factors Rap1, Ste12 and Cin5, respectively. Rap1 is known to bind incomplete or degenerate instances of the published motif (Table 2), and the new motif may confer additional specificity. The motif associated with Ste12 is the known binding site for the partner transcription factor Tec1, suggesting that Ste12 binding is strongly associated with its partner under the conditions examined (see below). Similarly, the new motif associated with Cin5 may be that of a partner transcription factor.
- Three new motifs are associated with the GO category for carbohydrate transport, fatty-acid oxidation and glycolysis–glycogenesis, respectively.
- Three new motifs are associated with an expression cluster (cluster 37) that includes many genes involved in energy metabolism and stress response. The cluster is also associated with a new genome-wide motif (motif 25).

Category-based motif discovery thus contributes a modest number of additional motifs beyond those found by genome-wide analysis.

Table 4 Additional new motifs discovered by category-based analysis

No.	Category*	Category-based motif†	Interpretation	Score‡
1	Exp.: cluster 37	YCCCTTAAA	New: cluster 37 (Msn2/4-like)	[8.5]
2	ChIP: <i>FHL1</i> in YPD	ATGTACGGATG	New: Rap1 alternate	[7.6]
3	GO: carbohydrate transport	GTITTTCCG	New: carbohydrate transport	[7.2]
4	GO: fatty acid beta-oxidation	TTAnnnCGG	New: fatty acid oxidation	[6.3]
5	GO: glycolysis/glyconeogenesis	TAGTGGAAGC	New: glycolysis/glycogenesis	[6.0]
6	Exp.: cluster 37	TCAGCC	New: cluster 37	[5.9]
7	Exp.: cluster 37	CGGnnnnCGG	New: cluster 37	[5.7]
8	ChIP: <i>CIN5</i> in YPD	GnTTAnnTnAGC	New: Cin5 alternate	[5.6]
9	ChIP: <i>STE12</i> in butanol	CATTCT	Known: Tec1	[5.4]

*Category used in search.
†Symbols as in Tables 2 and 3.
‡Category-based enrichment score.

Combinatorial control

Given a limited number of genome-wide regulatory motifs and a large number of transcriptionally regulated processes, fine-grained regulation may depend on combinatorial control. We sought to explore such combinatorial control by searching for motifs that occur in the same intergenic regions much more frequently than would be expected by chance.

In a single genome, few significant correlations are found. This is because functional instances of the motif are overwhelmed by a much larger number of random occurrences. Cross-species conservation greatly decreases this random noise and reveals biologically meaningful correlations. We describe a few examples here.

The Ste12 and Tec1 motifs show clear correlation, with about 20% of regions having a conserved occurrence of one also having a conserved occurrence of the other. This enrichment is not apparent when considering *S. cerevisiae* alone. We found that the genes that contain only the conserved Ste12 motif are enriched for those involved in mating and pheromone response, whereas those that contain conserved occurrences of both the Ste12 and Tec1 motifs are enriched for those involved in filamentous growth. These computational observations are consistent with recent work showing genome-wide evidence that Ste12 and Tec1 cooperate during starvation to induce filamentation-specific genes⁴⁹. We also found that genes that contain only conserved occurrences of the Tec1 motif are enriched for those involved in budding and cell polarity, suggesting that Tec1 has functions that do not require cooperative binding with Ste12.

About 60% of regions containing conserved motifs for the transcription factor Leu3 (which regulates branched-chain amino-acid biosynthesis) also contain conserved motifs for Gcn4 (a general factor regulating amino-acid biosynthesis, as well as many other processes). About 46% of regions containing conserved motifs for the transcription factor Met31 also contain conserved occurrences of Cbf1. In fact, Cbf1 (which is involved in DNA bending) is known to physically interact and cooperate with the MET regulatory complex. About 34% of regions containing a conserved Gal4 motif also contain a conserved Mig1 motif. In this case, the correlation reflects antagonistic interaction. Gal4 induces galactose metabolism genes in the presence of galactose, but Mig1 represses galactose metabolism in the presence of glucose.

Pairwise co-occurrence connects a group of five motifs: Msn2/4 (general stress response), Rlm1 (response to cell-wall stresses), Pdr1 (pleiotropic drug resistance), Tea1 (Ty element activator) and Tbf1 (Telomere-binding factor). This suggests a possible link between various stress responses and adaptive changes at the genome level⁵⁰.

Many additional correlations are seen among known and new motifs and can be used to construct comprehensive co-occurrence networks. These can provide information valuable in deciphering biological pathways in yeast.

Discussion

The goal of this paper is to explore the ability to extract a wide range of biological information from genome comparisons between related organisms. For *S. cerevisiae*, our results show that comparative genome analysis of a handful of related species has substantial power to identify genes, define gene structure, highlight rapid and slow evolutionary change, recognize regulatory elements and reveal combinatorial control of gene regulation. The power is comparable or superior to experimental analysis, in terms of sensitivity and precision.

In principle, the approach could be applied to any organism by selecting a suitable set of related species. The optimal choice of species depends on multiple considerations, largely related to the evolutionary tree connecting the species. These include the following: first, the branch length t between species should be short enough to permit orthologous sequence to be readily aligned. The

yeasts studied here differ by $t = 0.23$ – 0.55 substitutions per site and are readily aligned. The strong conservation of synteny allowed the unambiguous correspondence of the vast majority of genes.

Second, the total branch length of the tree should be large enough that non-functional sites will have undergone substantially more drift than functional sites, thereby providing an adequate degree of signal-to-noise enrichment. For this analysis, the multiple species studied provide a total branch length of 0.83 and a probability of nucleotide identity across all four species in non-coding regions of 49%. The signal-to-noise enrichment is thus about twofold (that is, $1/0.49$) for highly constrained nucleotides and correspondingly higher for composite features involving many nucleotides.

Third, the species should represent as narrow a taxon as possible, subject to the considerations above. Because the comparative analysis above seeks to identify genomic elements common to the species, it can explain only aspects of biology shared across the taxon. In the present case, the analysis identifies elements shared across *Saccharomyces sensu stricto*, a closely related set of species in which most of the genes and regulatory elements are shared.

What are the implications for the understanding of the human genome? The present study provides a good model for evolutionary distances (substitutions per site in intergenic regions) relevant to the study of the human. The sequence divergence between *S. cerevisiae* and the most distant relative *S. bayanus* (11% indels and 62% nucleotide identity in aligned positions) is similar to that between human and mouse (12% indels and 66% nucleotide identity in aligned positions).

An important difference between yeast and human is the inherent signal-to-noise ratio in the genome. Yeast has a high signal-to-noise ratio, with protein-coding regions comprising approximately 70% of the genome, and regulatory elements comprising perhaps about 15% of the intergenic regions. The human has a much lower signal-to-noise ratio, with the corresponding figures being perhaps ~2% and ~3%⁷. A lower signal-to-noise ratio must be offset by a higher signal-to-noise enrichment. Some enrichment can also be obtained by filtering out the repeat sequences that comprise half of the human genome. Greater enrichment can be accomplished by increasing the number of species studied, taking advantage both of nucleotide-level divergence and frequently occurring genomic deletion⁷.

Such considerations indicate that it should be possible to use comparative analysis, such as explored here for yeast, to identify directly many functional elements in the human genome that are common to mammals. More generally, comparative analysis offers a powerful and precise initial tool for interpreting genomes. □

Methods

Strain information

The sequence and annotation for *S. cerevisiae* reference strain S288C was obtained from the SGD website (<http://www-genome.stanford.edu/Saccharomyces/>) in May 2002. *Saccharomyces paradoxus* strain NRRL Y-17217, *S. mikatae* strain IFO1815, and *S. bayanus* strain MCYC623 were provided by E. Louis (University of Leicester). All sequenced isolates were diploid.

Determining one-to-one orthologous ORFs

For each species, we gathered the complete set of all predicted ORFs starting with a methionine and having a length of at least 50 amino acids. On the basis of blast hits, we connected predicted ORFs to *S. cerevisiae* ORFs in a bipartite graph, which we automatically separated into subgraphs by eliminating suboptimal edges in a series of steps. We first eliminated all edges that are less than 80% of the maximum-weight edge both in amino-acid identity and in length. On the basis of the resulting unambiguous matches, we built blocks of conserved gene order (synteny). We then preferentially kept matches within synteny blocks and resolved additional matches. We finally separated gene families into subfamilies by searching for best unambiguous subgraphs. We reported the connected components of the resulting graph as homology groups of unresolved genes.

Motif conservation score

To evaluate the motif conservation score (MCS) of a motif m of given length and degeneracy, we compared its conservation ratio to that of random patterns of the same length and degeneracy. We first computed the table F containing the relative frequencies of twofold and threefold degenerate bases, given the *S. cerevisiae* nucleotide frequencies. We then selected 20 random intergenic loci in *S. cerevisiae* and used the order of nucleotides at

each locus together with the order of degeneracy levels in m to construct 20 random motifs that were used as controls. We counted conserved and non-conserved instances of each control and computed r , the log-average of their conservation rates. We also counted the number of conserved and non-conserved intergenic instances of m , and evaluated the binomial probability p of observing the two counts, given r . We finally reported the MCS of the motif as a z-score corresponding to p , the number of standard deviations away from the mean of a normal distribution that corresponds to tail area p .

Conservation criteria

For CC1, we searched for mini-motifs that show a significant conservation in intergenic regions. For every mini-motif, we count i_c , the number of perfectly conserved intergenic instances in all four species, and i , the total number of intergenic instances in *S. cerevisiae*. We found that the two counts seem linearly related for most of the patterns (Fig. 7a), which can be attributed to a basal level of conservation r given the total evolutionary distance that separates the four species compared. We estimated the ratio r as the log-average of non-outlier instances of i_c/i within a control set of all motifs at a given gap size. We then calculated for every motif the binomial probability p of observing i_c successes out of i trials, given parameter r . We assigned a z-score S to every motif corresponding to probability p . Similar methods were used for computing CC2 and CC3 (see Supplementary Information).

Constructing full motifs

We extended each mini-motif selected by searching for surrounding bases that are preferentially conserved when the motif is conserved. We used an iterative approach adding at every iteration one base that maximally discriminates the neighbourhood of conserved motif instances from the neighbourhood of non-conserved motif instances. The added base was selected from 14 degenerate symbols of the IUB code. When no such symbol separated the conserved and non-conserved instances with significance above 3σ, we terminated the extension. Figure 7d shows the top-scoring mini-motif found in CC1 (panel 1), and the corresponding extension (panel 2). We found that many mini-motifs have the same or similar extensions, and we clustered the extended motifs hierarchically based on sequence similarity. We then computed a consensus sequence for every cluster of extended motifs, resulting in a smaller number of mega-motifs for each test (332 for CC1, 269 for CC2, and 285 for CC3). Panel 3 shows the first nine members of the top cluster in CC1, and the resulting mega-motif. Finally, we merged mega-motifs based on their co-occurrence in the same intergenic regions (panel 4). We computed a hypergeometric co-occurrence score between the intergenic regions containing each mega-motif and again collapsed these hierarchically. We computed a consensus for every cluster, and iterated the co-occurrence-based collapsing step (results not shown). We obtained fewer than 200 distinct genome-wide motifs, of which 72 have MCS ≥ 4 (Table 3).

Additional methods

See Supplementary Information for a complete description of methods for sequencing, assembly, annotation, nucleotide alignments, genomic rearrangements, ambiguous ORFs, reading frame conservation, gene boundaries, resequencing, intron finding, genome-wide motif discovery, motif conservation score, the three mini-motif conservation criteria, motif extension and collapsing, and category-based motif discovery.

Received 27 February; accepted 1 April 2003; doi:10.1038/nature01644.

1. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).
2. Kowalczyk, M., Mackiewicz, P., Gierlik, A., Dudek, M. R. & Cebrat, S. Total number of coding open reading frames in the yeast genome. *Yeast* **15**, 1031–1034 (1999).
3. Harrison, P. M., Kumar, A., Lang, N., Snyder, M. & Gerstein, M. A question of size: the eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* **30**, 1083–1090 (2002).
4. Velculescu, V. E. *et al.* Characterization of the yeast transcriptome. *Cell* **88**, 243–251 (1997).
5. Blandin, G. *et al.* Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Lett.* **487**, 31–36 (2000).
6. Wood, V., Rutherford, K. M., Ivens, A., Rajandream, M.-A. & Barrell, B. A Re-annotation of the *Saccharomyces cerevisiae* genome. *Comp. Funct. Genomics* **2**, 143–154 (2001).
7. International Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**, 520–562 (2002).
8. Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **2**, 28–36 (1994).
9. Tavazoie, S., Hughes, J. D., Campbell, M. J., Cho, R. J. & Church, G. M. Systematic determination of genetic network architecture. *Nature Genet.* **22**, 281–285 (1999).
10. Stormo, G. D. DNA binding sites: representation and discovery. *Bioinformatics* **16**, 16–23 (2000).
11. McGuire, A. M., Hughes, J. D. & Church, G. M. Conservation of DNA regulatory motifs and discovery of new motifs in microbial genomes. *Genome Res.* **10**, 744–757 (2000).
12. Loots, G. G. *et al.* Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. *Science* **288**, 136–140 (2000).
13. Pennacchio, L. A. & Rubin, E. M. Genomic strategies to identify mammalian regulatory sequences. *Nature Rev. Genet.* **2**, 100–109 (2001).
14. Oeltjen, J. C. *et al.* Large-scale comparative sequence analysis of the human and murine Bruton's tyrosine kinase loci reveals conserved regulatory domains. *Genome Res.* **7**, 315–329 (1997).
15. Clifton, P. F. *et al.* Surveying *Saccharomyces* genomes to identify functional elements by comparative DNA sequence analysis. *Genome Res.* **11**, 1175–1186 (2001).
16. Alm, R. A. *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397**, 176–180 (1999).
17. Carlton, J. M. *et al.* Genome sequence and comparative analysis of the model rodent malaria parasite *Plasmodium yoelii yoelii*. *Nature* **419**, 512–519 (2002).
18. Perrin, A. *et al.* Comparative genomics identifies the genetic islands that distinguish *Neisseria meningitidis*, the agent of cerebrospinal meningitis, from other *Neisseria* species. *Infect. Immun.* **70**, 7063–7072 (2002).

19. McClelland, M. *et al.* Comparison of the *Escherichia coli* K-12 genome with sampled genomes of a *Klebsiella pneumoniae* and three *salmonella enterica* serovars, *Typhimurium*, *Typhi* and *Paratyphi*. *Nucleic Acids Res.* **28**, 4974–4986 (2000).
20. Batzoglou, S. *et al.* ARACHNE: a whole-genome shotgun assembler. *Genome Res.* **12**, 177–189 (2002).
21. Gardner, M. J. *et al.* Genome sequence of the human malaria parasite *Plasmodium falciparum*. *Nature* **419**, 498–511 (2002).
22. Fischer, G., James, S. A., Roberts, I. N., Oliver, S. G. & Louis, E. J. Chromosomal evolution in *Saccharomyces*. *Nature* **405**, 451–454 (2000).
23. Dunham, M. J. *et al.* Characteristic genome rearrangements in experimental evolution of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **99**, 16144–16149 (2002).
24. Blanchette, M. & Tompa, M. Discovery of regulatory elements by a computational method for phylogenetic footprinting. *Genome Res.* **12**, 739–748 (2002).
25. Fischer, G., Neuveglise, C., Durrens, P., Gaillardin, C. & Dujon, B. Evolution of gene order in the genomes of two related yeast species. *Genome Res.* **11**, 2009–2019 (2001).
26. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
27. Bon, E. *et al.* Genomic exploration of the hemiascomycetous yeasts: 5. *Saccharomyces bayanus* var. *uvarum*. *FEBS Lett.* **487**, 37–41 (2000).
28. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
29. Dujon, B. *et al.* Complete DNA sequence of yeast chromosome XI. *Nature* **369**, 371–378 (1994).
30. Sharp, P. M. & Li, W. H. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
31. Clark, T. A., Sugnet, C. W. & Ares, M. Jr Genome-wide analysis of mRNA processing in yeast using splicing-specific microarrays. *Science* **296**, 907–910 (2002).
32. Hurst, L. D. The Ka/Ks ratio: diagnosing the form of sequence evolution. *Trends Genet.* **18**, 486 (2002).
33. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
34. True, H. L. & Lindquist, S. L. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* **407**, 477–483 (2000).
35. Koufopanou, V., Goddard, M. R. & Burt, A. Adaptation for horizontal transfer in a homing endonuclease. *Mol. Biol. Evol.* **19**, 239–246 (2002).
36. Haber, J. E. Mating-type gene switching in *Saccharomyces cerevisiae*. *Annu. Rev. Genet.* **32**, 561–599 (1998).
37. Hampson, S., Kibler, D. & Baldi, P. Distribution patterns of over-represented k-mers in non-coding yeast DNA. *Bioinformatics* **18**, 513–528 (2002).
38. McCue, L. *et al.* Phylogenetic footprinting of transcription factor binding sites in proteobacterial genomes. *Nucleic Acids Res.* **29**, 774–782 (2001).
39. Gelfand, M. S., Koonin, E. V. & Mironov, A. A. Prediction of transcription regulatory sites in Archaea by a comparative genomic approach. *Nucleic Acids Res.* **28**, 695–705 (2000).
40. Keegan, L., Gill, G. & Ptashne, M. Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231**, 699–704 (1986).
41. Zhu, J. & Zhang, M. Q. SCPD: a promoter database of the yeast *Saccharomyces cerevisiae*. *Bioinformatics* **15**, 607–611 (1999).
42. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **27**, 44–48 (1999).
43. Dwight, S. S. *et al.* *Saccharomyces* Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res.* **30**, 69–72 (2002).
44. Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
45. Gasch, A. P. & Eisen, M. B. Exploring the conditional coregulation of yeast gene expression through fuzzy k-means clustering. *Genome Biol.* **3** RESEARCH0059 (2002).
46. Mosley, A. L., Lakshmanan, J., Aryal, B. K. & Ozcan, S. Glucose-mediated phosphorylation converts the transcription factor Rgt1 from a repressor to an activator. *J. Biol. Chem.* **278**, 10322–10327 (2003).
47. Lindgren, A. *et al.* The pachytene checkpoint in *Saccharomyces cerevisiae* requires the Sum1 transcriptional repressor. *EMBO J.* **19**, 6489–6497 (2000).
48. Jacobs Anderson, J. S. & Parker, R. Computational identification of cis-acting elements affecting post-transcriptional control of gene expression in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **28**, 1604–1617 (2000).
49. Zeitlinger, J. *et al.* Program-specific distribution of a transcription factor dependent on partner transcription factor and MAPK signaling. *Cell* **113**, 395–404 (2003).
50. Morillon, A., Springer, M. & Lesage, P. Activation of the Kss1 invasive-filamentous growth pathway induces Tyl1 transcription and retrotransposition in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **20**, 5766–5776 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. Botstein, M. Cherry, K. Dolinski, D. Fisk, S. Weng and other members of the *Saccharomyces* Genome Database staff for assistance with SGD, for making our data available to the community through SGD, and for discussions; J. Butler, S. Calvo, J. Galagan, D. Jaffe, J. Lehar and L. Jun Ma for technical advice and discussions; the staff of the Whitehead/MIT Center for Genome Research Sequencing Center who generated the shotgun sequence from the three yeast species; T. Lee, N. Rinaldi, R. Young and J. Zeitlinger for sharing data about chromatin immunoprecipitation experiments and for discussions; M. Eisen and A. Gasch for sharing information about gene expression clusters and for discussions; E. Louis and I. Roberts for providing yeast strains and discussions; B. Berger, G. Fink, D. Gifford, S. Lindquist and H. True-Krobb for discussions; and L. Gaffney for assistance with figures.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.K. (manoli@mit.edu) or E.S.L. (lander@wi.mit.edu). GenBank accession numbers are AABY01000000 for *S. paradoxus*, AABZ01000000 for *S. mikatae*, and AAC01000000 for *S. bayanus*.

Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells

Julie Lessard* & Guy Sauvageau*†‡

*Laboratory of Molecular Genetics of Hemopoietic Stem Cells, Clinical Research Institute of Montreal, 110 Pine Avenue West, Montreal, Quebec, Canada, H2W 1R7
†Department of Medicine and ‡Division of Hematology, Hospital Maisonneuve-Rosemont, Montreal, Montreal University, Montreal, Quebec, Canada, H3C 3J7

An emerging concept in the field of cancer biology is that a rare population of 'tumour stem cells' exists among the heterogeneous group of cells that constitute a tumour. This concept, best described with human leukaemia, indicates that stem cell function (whether normal or neoplastic) might be defined by a common set of critical genes. Here we show that the *Polycomb* group gene *Bmi-1* has a key role in regulating the proliferative activity of normal stem and progenitor cells. Most importantly, we provide evidence that the proliferative potential of leukaemic stem and progenitor cells lacking *Bmi-1* is compromised because they eventually undergo proliferation arrest and show signs of differentiation and apoptosis, leading to transplant failure of the leukaemia. Complementation studies showed that *Bmi-1* completely rescues these proliferative defects. These studies therefore indicate that *Bmi-1* has an essential role in regulating the proliferative activity of both normal and leukaemic stem cells.

In humans, the concept of 'tumour stem cell' is best described in acute myeloid leukaemias (AMLs) in which the majority of the leukaemic cells (blasts) display a limited proliferative capacity and must therefore be constantly regenerated by the rare 'leukaemic stem cell' (L-HSC) population, which is itself capable of extensive self-renewal^{1–3} (see also Supplementary Fig. 1a). Interestingly, these L-HSCs share phenotypical similarities (but also some differences, such as CD123 expression^{4,5}) with normal haematopoietic stem cells (HSCs)¹.

Several arguments suggest that the *Polycomb* group (*PcG*) gene *Bmi-1* is a good candidate for regulating the proliferative activity of both normal and leukaemic HSCs. First, *Bmi-1* expression is restricted to primitive bone marrow cells in both humans and mice^{6–8} and is expressed in all myeloid leukaemias analysed so far⁶ including the L-HSC-enriched CD34⁺ fraction (Supplementary Fig. 1b). Second, homozygosity for a null allele of *Bmi-1* causes a profound and progressive failure of the entire haematopoietic system in mice that is probably responsible for their death in early adulthood or before⁹. This progressive and fatal depletion of all blood cells (including primitive progenitors⁷) indicates that *Bmi-1* is not required for the generation of HSCs derived from fetal liver but essential for adult HSCs. Third, in addition to this presumed function in HSCs, *Bmi-1* has a critical and dose-dependent role in regulating the proliferative output of primitive progenitors derived from bone marrow⁷.

Here we provide evidence demonstrating that *Bmi-1* is dispensable for the generation of stem and progenitor cells derived from fetal liver, but is absolutely necessary for their full proliferative potential. Most interestingly, our studies indicate that this function of *Bmi-1* is preserved in leukaemic stem and progenitor cells and therefore provide a molecular basis for the concept that common genetic determinants regulate normal and leukaemic stem cells.

Function of *Bmi-1* in stem and progenitor cells

Bmi-1^{−/−} E14.5 fetal liver cells grown for 4 days *in vitro* (Fig. 1a) present a 27-fold reduction in their frequency of myeloid progenitors when compared to *Bmi-1*^{+/+} controls (Fig. 1c, left panel). This progenitor cell defect was quantitatively and qualitatively rescued by the retroviral transduction (94–100% gene transfer) of *Bmi-1* in

these cells (see Fig. 1c, right panel), indicating that *Bmi-1* is dispensable for the generation of myeloid progenitors derived from fetal liver but absolutely essential for their full proliferative activity.

On the basis of these results, it can be inferred that a similar situation might occur at the long-term repopulating cell (LTRC) level. An inability of bone marrow and E14.5 fetal-liver-derived *Bmi-1*^{−/−} cells to contribute to long-term haematopoiesis was

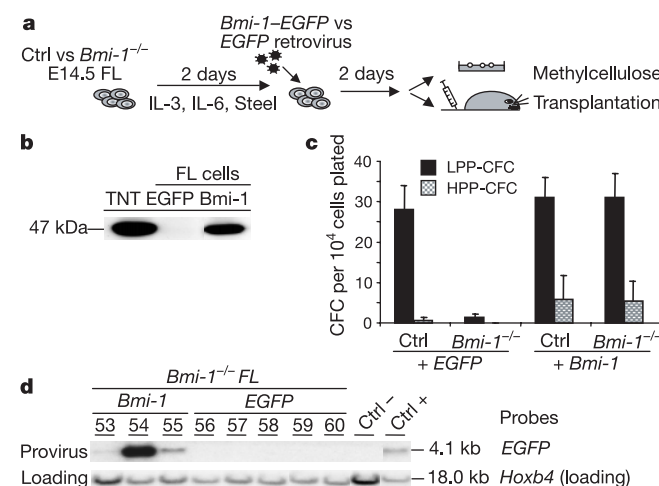


Figure 1 *Bmi-1* regulates the proliferative potential of embryonic day (E)14.5 fetal liver (FL)-derived haematopoietic cells. **a**, Overview of the experimental strategy. Ctrl, control; IL, interleukin. **b**, Western blot analysis of total cell extracts isolated from fetal liver cells engineered to express *Bmi-1* (lane 3) or not (lane 2). Lane 1, transcribed and translated (TNT) *Bmi-1* protein. **c**, *Bmi-1*^{−/−} and *Bmi-1*^{+/+} fetal liver cells engineered to express EGFP or *Bmi-1*-EGFP were assayed for their content in colony-forming cells (CFC). HPP-CFC and LPP-CFC indicate CFC with high and low proliferative potential. **d**, Southern blot analysis to detect the integrated proviruses in DNA (digested with *KpnI*) from the bone marrow of recipients of *Bmi-1*^{−/−} fetal liver cells transduced with *Bmi-1* (lanes 1–3) or with EGFP only (lanes 4–8). kb, kilobases.

recently shown in reconstitution experiments, suggesting that *Bmi-1* plays a key role in self-renewal/proliferation of primitive haematopoietic cells²⁴. Similarly, an experiment performed by our group (with the same group of cells as described in Fig. 1c, transplanted at about one LTRC per sublethally irradiated recipient) confirmed the presence of roughly normal numbers of cells with long-term repopulating potential in the fetal liver of *Bmi-1*^{-/-} mice but the detection of these cells, at 16 weeks after transplantation, was strictly dependent on the retroviral expression of *Bmi-1* (Fig. 1d; three of three recipients were reconstituted with *Bmi-1*-transduced cells, whereas no reconstitution was observed in five of five recipients of a similar number of cells transduced with enhanced green fluorescent protein (EGFP)). As might be expected from previous experiments in which near-limiting numbers of stem cells are transplanted, the proportion of EGFP-expressing lymphoid and myeloid cells was at best 5.6% in the peripheral blood of these mice (nos 53–55 in Fig. 1d). The long-term (16 weeks) and pluripotent potential of these cells confirmed that the rescue was occurring at the HSC level.

Together, these complementation studies confirmed the critical function of *Bmi-1* once stem and progenitor cells have reached a given developmental stage. Considering the striking parallel between the organization of normal and leukaemic haematopoiesis and the knowledge that *Bmi-1* is expressed in all primary leukaemia and leukaemic cell lines analysed so far (see the introduction and Supplementary Fig. 1b), we sought to investigate the importance of *Bmi-1* in the biology of leukaemic stem and progenitor cells.

Generation of leukaemia from *Bmi-1*^{-/-} fetal liver cells

To determine whether *Bmi-1* might also be essential in leukaemic haematopoiesis as it is in the normal situation, we first investigated whether stem or progenitor cells lacking *Bmi-1* could be trans-

formed by a set of complementary oncogenes, and whether these leukaemic cells could self-renew and proliferate sufficiently to induce AML in mice. To achieve this, *Bmi-1*^{-/-} fetal liver cells were infected with a retrovirus containing both the *Hoxa9* and *Meis1a* genes, and transduced cells were transplanted into sublethally irradiated syngeneic hosts (Fig. 2a). The selection of the *Hoxa9* and *Meis1* collaborating oncogenes was based on several criteria: first, these genes generate a well-characterized myeloid leukaemia in mice that displays a similar cellular hierarchy (namely L-HSCs and 'mature' blasts) to that observed in human patients¹⁰; second, these two oncogenes are sufficient to induce full transformation of primary haematopoietic cells without the need for additional genetic events¹⁰; third, the HSCs, or a close progeny, are the target for the transformation induced by these genes (ref. 10 and U. Thorsteinsdottir, P. Austin and G.S., unpublished data); and last, the *Hoxa9* oncogene is involved in myeloid leukaemia in humans^{11,12}.

Regardless of the genotype (that is, *Bmi-1*^{+/-} or *Bmi-1*^{-/-}), all recipients of *Hoxa9*–*Meis1*-transduced fetal liver cells developed AML within a similar latency period (Fig. 2b). The simultaneous overexpression of *Bmi-1* did not accelerate the occurrence of the

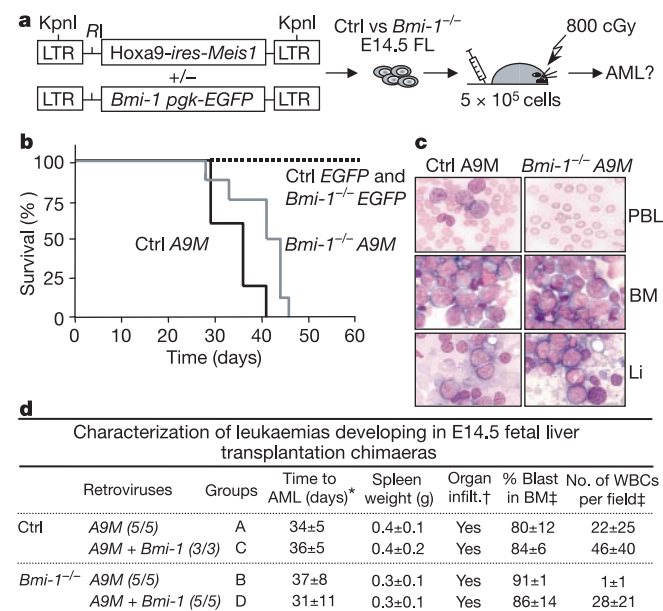


Figure 2 *Bmi-1* is dispensable for the generation of AML in primary recipients.

a, Summary of the experimental strategy and retroviruses used. Ctrl, control; FL, fetal liver; LTR, long terminal repeat. **b**, Survival graph of mice reconstituted with control or *Bmi-1*^{-/-} E14.5 fetal liver cells engineered to express either *Hoxa9*–*Meis1* (A9M) or EGFP. **c**, Cytological preparations of peripheral blood leukocytes (PBL), bone marrow (BM) and liver (Li) of representative leukaemic mice from each group shown in **b**. **d**, Major characteristics of the leukaemia that occurred in each cohort. * Mean ± SD; † organs infiltrated were lung, liver, kidney; ‡ *n* = 200 cells were counted from three representative mice. WBCs, white blood cells.

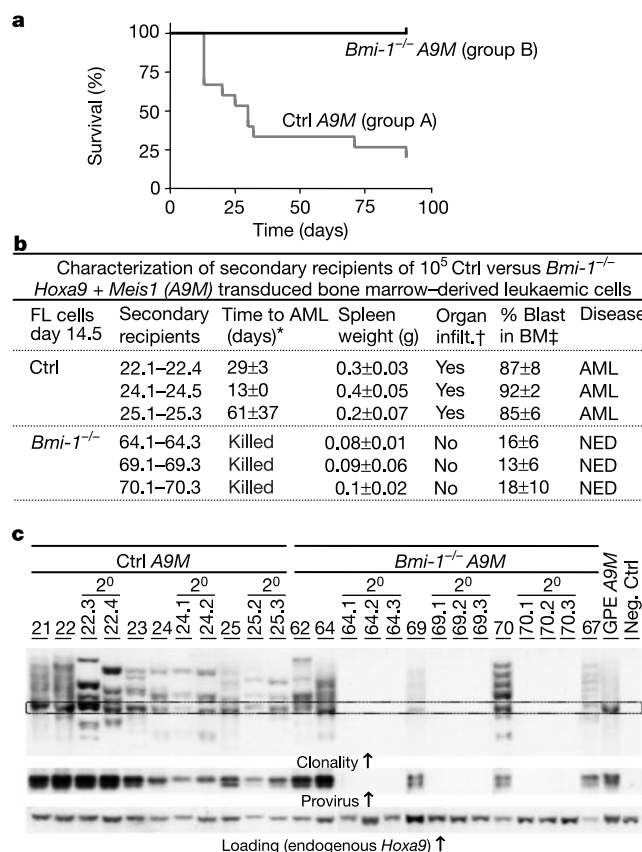


Figure 3 *Bmi-1* is essential for the 'transplantability' of AML into secondary hosts.

a, Survival graph of secondary recipients of control (Ctrl) and *Bmi-1*^{-/-} leukaemic cells (see Fig. 2). **b**, Characterization of secondary recipients shown in **a**. Abbreviations and details are as in Fig. 2d. FL, fetal liver; NED, no evidence of disease. **c**, Southern blot analyses of DNA isolated from the bone marrow of primary and secondary recipients detailed in Fig. 2d and 3b. 2⁰, Secondary recipients. Top panel versus middle and lower panels show *EcoRI* versus *KpnI* restriction, respectively. The probe was *Hoxa9*. The bands enclosed in the dotted box in the top panel correspond to a specific fragment of proviral *Hoxa9*-derived DNA that does not represent clonality (possibly representing 'star' or inadequate activity). GPE A9M, viral producers of *Hoxa9*–*Meis1* retrovirus; Neg. ctrl, GPE cells only.

leukaemia in all groups analysed, nor did it change their phenotype or clinical presentation (except for one parameter; see below and Fig. 2c and d). The leukaemic mice from all groups had a very similar and aggressive disease (see Fig. 2c and d for a detailed description). When analysed by fluorescence-activated cell sorting (FACS), the leukaemic blasts were strongly positive (more than 96%) for the Mac-1 myeloid cell surface antigen and 13–33% co-expressed Gr-1 (data not shown).

Although the *Bmi-1* status of the transformed fetal liver cells did not affect the phenotypic and clinical characteristics of the leukaemias, the presence of *Bmi-1* seemed to be important for the capacity for large numbers of leukaemic cells to accumulate in the peripheral blood (Fig. 2d, group A versus group B, eighth column). This phenotype was dependent on *Bmi-1* because it was complemented by the retroviral introduction of *Bmi-1* in *Bmi-1*^{-/-} fetal liver cells (Fig. 2d, group A versus group D, eighth column).

A northern blot analysis of total RNA isolated from the leukaemic cells confirmed high expression levels of retrovirally derived mRNAs (data not shown). DNA analysis of proviral integration sites in both wild-type and *Bmi-1*^{-/-} leukaemic cells demonstrated that the *Hoxa9-Meis1*-induced AMLs were derived from two to four leukaemic clones (data not shown). Because a leukaemic mouse contains at least 10⁹ leukaemic cells, this indicates that *Bmi-1*^{-/-} leukaemic cells displayed a substantial proliferative capacity *in vivo*. However, the lower leukaemic cell numbers in the peripheral blood of recipients of *Bmi-1*^{-/-} cells (Fig. 2d, group B) could indicate a progressive exhaustion of the *Bmi-1*^{-/-} L-HSCs.

Bmi-1^{-/-} AMLs do not repopulate secondary recipients

To evaluate a possible exhaustion of *Bmi-1*^{-/-} L-HSCs, a dose of 10⁵ leukaemic cells derived from the bone marrow of three different primary recipients from group A (control *Hoxa9-Meis1*) and group B (*Bmi-1*^{-/-} *Hoxa9-Meis1*) were transplanted into syngeneic hosts (*n* = 5 secondary recipients per primary leukaemic mouse). Mice were then monitored over time for the development of AML. Strikingly, *Bmi-1* seemed to be essential for the maintenance of *Hoxa9-Meis1* L-HSCs *in vivo* because the leukaemias generated in *Bmi-1*^{-/-} cells were not maintained on transplantation into

secondary hosts (Fig. 3a, group B). However, animals bearing control *Hoxa9-Meis1* transplanted tumours developed AML within 30 ± 25 days after transplantation (Fig. 3a and b, group A). Clonal analyses of proviral integration sites indicated these leukaemias originated from several clones (Fig. 3c, and data not shown).

As mentioned above, none of the secondary recipients of *Bmi-1*^{-/-} *Hoxa9-Meis1* leukaemic bone marrow cells developed AML (Fig. 3a), and no leukaemic cells were detected in long-term recipients analysed at 4 and 11 months after transplantation (Fig. 3b and c, and data not shown, respectively).

This difference in the 'transplantability' between *Bmi-1*^{-/-} and control AMLs is particularly significant considering that all (*n* = 3) of the secondary recipients receiving as few as 1,000 leukaemic cells from the control 'group A' developed AML within 27 ± 9 days of transplantation (data not shown). In all cases, the invasive behaviour of the control *Hoxa9-Meis1* tumours was reproduced after transplantation, and the morphology and phenotype of the leukaemias were indistinguishable (data not shown). Interestingly, leukaemias developing in two of three primary recipients of *Bmi-1*^{-/-} *Hoxa9-Meis1* fetal liver cells co-infected with *Bmi-1* (group D) were readily transferred to secondary recipients within a similar timeframe to that of control AMLs (group A) (data not shown). This suggests that the transplantation of the leukaemic phenotype is dependent on the continuous expression of *Bmi-1*.

In vitro properties of *Bmi-1*^{-/-} leukaemic cells

To explain the apparent exhaustion of the *Bmi-1*^{-/-} leukaemic clones *in vivo*, we monitored their survival and growth characteristics *in vitro*. After 10 days of growth, the cellularity of cultures initiated with *Bmi-1*^{-/-} leukaemic cells was decreased 15(±4)-fold relative to controls (namely *Bmi-1*^{+/+}; data not shown). Supporting these findings, a direct comparison between control and *Bmi-1*^{-/-} leukaemic cells indicated that the latter had the following properties: first, they displayed accumulation in G1 (Fig. 4a and b) and reduction in S phase (Fig. 4c); second, they underwent much greater morphological changes consistent with differentiation to butyrate-positive, adherent-layer-forming macrophages; third, they showed

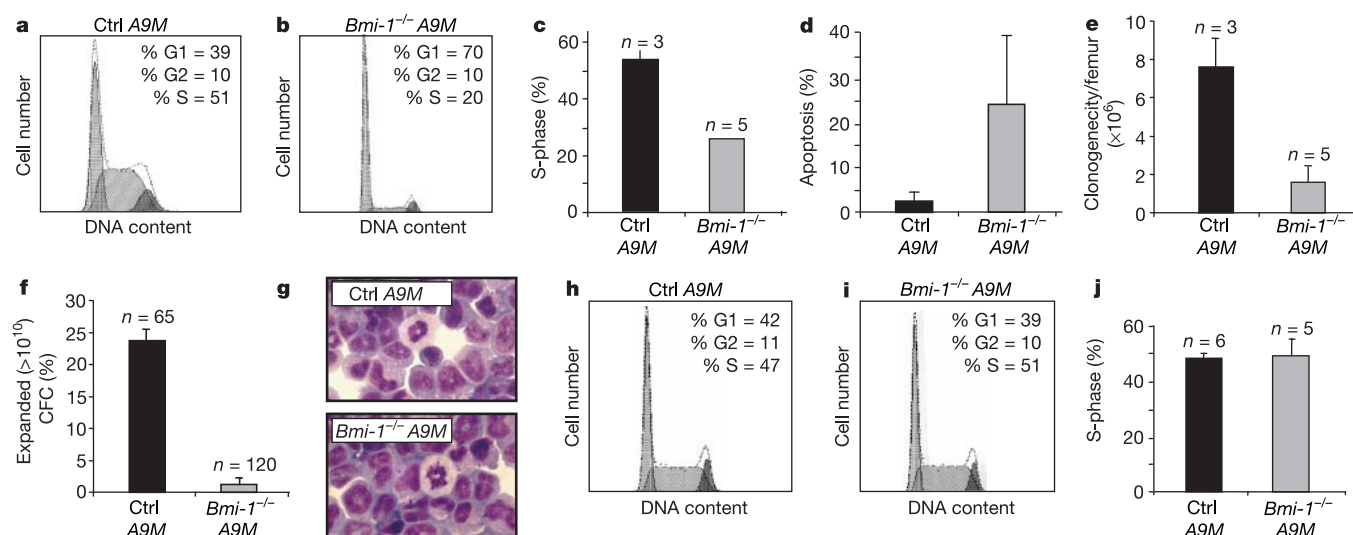


Figure 4 Characterization of control and *Bmi-1*^{-/-} leukaemic cells *in vitro*. **a–c**, Cell cycle analysis of *Bmi-1*^{+/+} (**a**) and *Bmi-1*^{-/-} leukaemic cells (**b**) performed as described²². **c**, Average proportion of cells from **a** and **b** in S-phase. Ctrl, control. **d**, Apoptotic index determined by the percentage of cells positive for both propidium iodine and annexin V. **e**, Absolute number of leukaemic colony-forming cells (CFCs) per

femur in primary recipients. **f**, Percentage of CFCs able to generate HPCs (highly proliferative clones). **g**, Wright staining of representative HPCs from **f**. **h**, Cell cycle analysis of *Bmi-1*^{+/+} (**h**) and *Bmi-1*^{-/-} (**i**) exponentially growing HPCs. **j**, Proportion of cells in S-phase in exponentially growing HPCs. For **c–f** and **j**, results are presented as means ± s.d.

an increased proportion of apoptotic cells (Fig. 4d); and last, they were less efficient in generating leukaemic colony-forming cells (Fig. 4e).

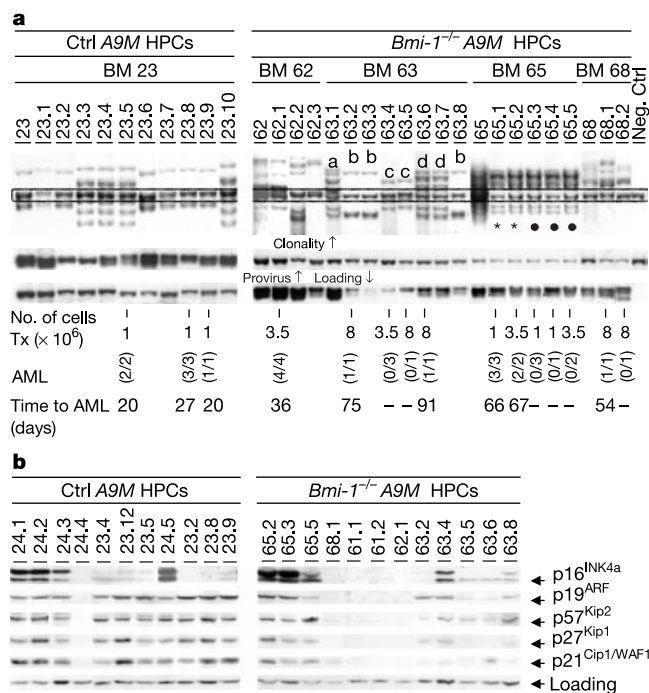
Derivation of clones from *Bmi-1*^{-/-} AMLs

To investigate further the impairment in proliferation of *Bmi-1*^{-/-} leukaemic blasts, attempts were made to derive highly proliferative clones (HPCs) from cultures initiated with *Bmi-1*^{-/-} and *Bmi-1*^{+/+} leukaemic cells (Supplementary Information). Only 1% of colonies from the *Bmi-1*^{-/-} cultures could generate more than 10¹⁰ cells (our minimal definition of HPC), whereas HPCs could be derived at a frequency of 24% in control cultures (Fig. 4f). In contrast to controls, all *Bmi-1*^{-/-} HPCs initially presented the same characteristics as those described in the previous section for leukaemic bone marrow cells (namely reduced proliferation, monocytic differentiation and increased apoptotic index). However, within about 2 weeks of culture, the *Bmi-1*^{-/-} HPCs were similar to controls (Fig. 4g–j).

Karyotype analysis (G-banding) performed on several independent control (group A, *n* = 3) and *Bmi-1*^{-/-} (group B, *n* = 8) HPCs revealed a near-diploid content (containing 38–40 chromosomes) and the absence of recurrent structural chromosomal aberrations, indicating that the derivation of *Bmi-1*^{-/-} HPCs *in vitro* involved neither overt genomic instability nor recurrent chromosomal lesions (data not shown). When analysed by FACS, about 100% of cells from HPCs from both groups A (*n* = 12) and B (*n* = 13) expressed the myeloid antigen Mac-1, whereas the expression of Gr-1 varied extensively (data not shown). Clonal analysis performed on several of the control (*n* = 37) and *Bmi-1*^{-/-}

(*n* = 25) HPCs revealed the presence of several integrations of an intact *Hoxa9*–*IRES*–*Meis1* provirus, where IRES stands for internal ribosome entry site (Fig. 5a, and data not shown).

The low frequency (1%) of clonal derivation from *Bmi-1*^{-/-} AMLs suggested either that there was a marked decrease in their generation *in vivo* and/or that genetic or epigenetic events were required for their emergence (clonal evolution). On the basis that the *Ink4a*–*Arf* locus is a critical genuine target *in vivo* for *Bmi-1* in the regulation of haematopoietic cell proliferation¹³, and increased p19^{ARF} levels induce a p53-mediated checkpoint response in G1 (refs 14, 15), we verified whether inactivation of Cdkn2a (p16^{INK4a}), Cdkn2b (p19^{ARF}) and other regulator(s) of the G1 cell cycle checkpoint was present in *Bmi-1*^{-/-} *Hoxa9*–*Meis1* HPCs. Most of the control *Hoxa9*–*Meis1* (group A) HPCs analysed presented normal levels of most of the G1 cyclin-dependent kinase inhibitors (CKIs) tested, including p19^{ARF}, p57^{Kip2}, p27^{Kip1} and p21^{Cip1} (Fig. 5b, left panel). In sharp contrast, most of the *Bmi-1*^{-/-} HPCs (group B) analysed showed low levels or complete loss of expression of several of these CKIs (*n* = 11 of 14 HPCs; Fig. 5b, right panel, and data not shown). Together with the initial delay (2 weeks) in proliferation of each of these *Bmi-1*^{-/-} HPCs (see above), these results indicate that clonal evolution, through



epigenetic or genetic loss of CKI expression, might represent the 'driving force' that allowed the derivation of HPCs from *Bmi-1*^{-/-} AMLs *in vitro*.

Bmi-1^{-/-} HPCs are weakly leukaemogenic

The ability of these HPCs to induce leukaemia *in vivo* was investigated by transplanting between 10⁶ and 8 × 10⁶ cells per clone into sublethally irradiated syngeneic recipients (Fig. 6a). Strikingly, all recipients of control *Hoxa9-Meis1* HPCs (*n* = 6, one to four mice per group) succumbed to AML within 16 ± 6 days of transplantation. However, all of the HPCs that could be derived from *Bmi-1*^{-/-} *Hoxa9-Meis1* AMLs either did not induce AML (6 of 12; Fig. 5a and 6b) or induced the disease after a significant delay (see below). DNA analysis of proviral integration sites from selected HPCs indicated that clones such as 68.2 or 65.3–65.5, which represented dominant leukaemic clones in primary recipients, failed to induce AML *in vivo*, whereas clones 65.1 and 65.2 (see asterisk in Fig. 5a), which could transmit AML to secondary recipients after long latencies (66–67 days), had the same retroviral integration profile as clones 65.3–65.5 (see filled circles in Fig. 5a), which could not.

Rescue of *Bmi-1*^{-/-} HPCs' AML-inducing capacity

To distinguish between clonal evolution and selection as the mechanism responsible for leukaemic progression in certain *Bmi-1*^{-/-} HPCs, we engineered the retroviral expression of *Bmi-1* or *EGFP* (control) in three HPCs (two non-leukaemic identified as HPC-63.5 and HPC-65.4, and one poorly leukaemic identified as HPC-63.2 in Fig. 6c; see also Fig. 5a and b) and transplanted these cells immediately after gene transfer, at a dose of 3.5 × 10⁶ cells, into irradiated syngeneic recipients (*n* = 2 or 3 mice per group, except HPC-63.2 plus *EGFP* (*n* = 1)). The three clones engineered to express *Bmi-1* readily induced AML within short latencies, whereas no leukaemias were detected until day 85 in recipients of similar numbers of cells from the same clones infected with the control *EGFP* virus (Fig. 6c). The leukaemias generated by the three HPCs expressing *Bmi-1* were highly polyclonal (see clonality in lanes 5–17 in Fig. 6d) and contained an intact *Bmi-1* provirus. The contribution of these *Bmi-1*^{-/-} clones expressing *Bmi-1* to AML (chimaerism) was evaluated by FACS analysis for EGFP expression (the *Bmi-1* virus contains an *EGFP* reporter gene) (Fig. 6d) and by Southern blot analysis, by assessing the ratio of the inactivated *Bmi-1* allele (donor-derived) relative to the wild-type *Bmi-1* allele (host-derived) in various haematopoietic organs of the leukaemic animals (Fig. 6d, bottom panel). Donor-derived cells corresponded to 25–95% of the total haematopoietic cell population, confirming infiltration of the neoplastic cells into all of these organs (Fig. 6d). The demonstration that *Bmi-1* was sufficient to fully rescue the leukaemogenic potential of *Bmi-1*^{-/-} HPCs further suggests that clonal evolution, rather than selection, is the underlying mechanism responsible for leukaemic progression in *Bmi-1*^{-/-} HPCs.

In summary, the findings reported here reinforce the notion of a structure in leukaemic hierarchy where 'stemness' would be conferred by the continual expression of *Bmi-1* (Supplementary Fig. 2). Importantly, it was possible under conditions of stimulation with high concentrations of growth factors *in vitro* to generate non-leukaemogenic *Bmi-1*^{-/-} clones with an impaired expression of p19^{ARF} and p16^{INK4a}, both of which are known functional targets of *Bmi-1* (ref. 13). Retroviral introduction of *Bmi-1* into these clones deficient in p19^{ARF} and p16^{INK4a} (*n* = 3) readily rescued their tumorigenic properties, suggesting that *Bmi-1* has one or more additional functions in L-HSCs besides repression of these CKIs. The expression of currently known regulators of early haematopoiesis (such as *tal-1/SCL* (ref. 16) and *Hoxb4* (ref. 17)) was not altered by the status of *Bmi-1* in leukaemic and non-leukaemic clones (Supplementary Fig. 3).

The apparent difficulties of bypassing the requirement for *Bmi-1* in leukaemic stem/progenitor cells *in vivo* suggests that adroit molecular targeting of *Bmi-1* in leukaemic stem/progenitor cells might have potent and specific therapeutic effects. Interestingly, *BMI-1* expression was recently reported in several cases of human non-small-cell lung cancer¹⁸, breast cancer cell lines¹⁹ and immortalized mammary epithelial cells (MECs)¹⁹. It will therefore be interesting to determine whether the findings reported here might also extend to other types of 'cancer stem cells'. □

Methods

Mice and genotyping

The *Bmi-1*^{-/-} mice (129Ola/FVB/N hybrid background) used in these studies⁹ had been backcrossed 10–12 times in the C57Bl/6J background.

Generation of haematopoietic chimaeras and viruses

Details of the construction of the *MSCV-Hoxa9-IRES-Meis1a* bicistronic retroviral vector and the *MSCV-Bmi-1-pgk-EGFP* vector are available from the authors on request. Production of vesicular stomatitis virus-pseudotyped retroviruses, infection of haematopoietic cells and transplantation into mice were performed as described²⁰.

Cellular analyses and progenitor determination

Analyses of progenitor cell numbers in methylcellulose cultures were done as reported^{10,21}. Cell cycle and flow cytometric analyses were performed as described^{21,22}.

Generation, amplification and analysis of complementary DNA

Semi-quantitative PCR with reverse transcription was performed with a previously described method that amplifies all the cDNAs derived from as few as 1–1,000 cells²³. The amplified (total) cDNAs were then transferred to a nylon membrane and hybridized to specific probes as indicated. See Supplementary Table 1 for a detailed description of each probe used.

Southern, northern and western analyses

All of these were essentially done as described^{10,22}. Information on the antibodies used in these studies is available from the authors on request.

Received 30 January; accepted 5 March 2003; doi:10.1038/nature01572.

Published online 20 April 2003.

- Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature Med.* **3**, 730–737 (1997).
- Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature* **367**, 645–648 (1994).
- Appelbaum, F. R., Rowe, J. M., Radich, J. & Dick, J. E. Acute myeloid leukemia. *Hematology (Am. Soc. Hematol. Educ. Program)*, 62–86 (2001).
- Jordan, C. T. *et al.* The interleukin-3 receptor alpha chain is a unique marker for human acute myelogenous leukemia stem cells. *Leukemia* **14**, 1777–1784 (2000).
- Testa, U. *et al.* Elevated expression of IL-3Rα in acute myelogenous leukemia is associated with enhanced blast proliferation, increased cellularity, and poor prognosis. *Blood* **100**, 2980–2988 (2002).
- Lessard, J., Baban, S. & Sauvageau, G. Stage-specific expression of polycomb group genes in human bone marrow cells. *Blood* **91**, 1216–1224 (1998).
- Lessard, J. *et al.* Functional antagonism of the Polycomb-Group genes *eed* and *Bmi1* in hemopoietic cell proliferation. *Genes Dev.* **13**, 2691–2703 (1999).
- Park, I. K. *et al.* Differential gene expression profiling of adult murine hematopoietic stem cells. *Blood* **99**, 488–498 (2002).
- van der Lugt, N. M. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev.* **8**, 757–759 (1994).
- Kroon, E. *et al.* *Hoxa9* transforms primary bone marrow cells through specific collaboration with *Meis1a* but not *Pbx1b*. *EMBO J.* **17**, 3714–3725 (1998).
- Borrow, J. *et al.* The t(7;11)(p15;p15) translocation in acute myeloid leukaemia fuses the genes for nucleoporin NUP98 and class I homeoprotein HOXA9. *Nature Genet.* **12**, 159–167 (1996).
- Nakamura, T. *et al.* Fusion of the nucleoporin gene NUP98 to HOXA9 by the chromosome translocation t(7;11)(p15;p15) in human myeloid leukaemia. *Nature Genet.* **12**, 154–158 (1996).
- Jacobs, J. J., Kieboom, K., Marino, S., DePinho, R. A. & van Lohuizen, M. The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* **397**, 164–168 (1999).
- Weber, H. O., Samuel, T., Rauch, P. & Funk, J. O. Human p14(ARF)-mediated cell cycle arrest strictly depends on intact p53 signaling pathways. *Oncogene* **21**, 3207–3212 (2002).
- Kamijo, T. *et al.* Functional and physical interactions of the ARF tumour suppressor with p53 and Mdm2. *Proc. Natl Acad. Sci. USA* **95**, 8292–8297 (1998).
- Porcher, C. *et al.* The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**, 47–57 (1996).
- Sauvageau, G. *et al.* Overexpression of HOXB4 in hematopoietic cells causes the selective expansion of more primitive populations *in vitro* and *in vivo*. *Genes Dev.* **9**, 1753–1765 (1995).
- Vonlanthen, S. *et al.* The *bmi-1* oncoprotein is differentially expressed in non-small cell lung cancer and correlates with INK4A-ARF locus expression. *Br. J. Cancer* **84**, 1372–1376 (2001).

19. Dimri, G. P. *et al.* The Bmi-1 oncogene induces telomerase activity and immortalizes human mammary epithelial cells. *Cancer Res.* **62**, 4736–4745 (2002).
20. Thorsteinsdottir, U. *et al.* Overexpression of the myeloid leukemia-associated Hoxa9 gene in bone marrow cells induces stem cell expansion. *Blood* **99**, 121–129 (2002).
21. Sauvageau, G. *et al.* Overexpression of HOXB3 in hematopoietic cells causes defective lymphoid development and progressive myeloproliferation. *Immunity* **6**, 13–22 (1997).
22. Kros, J. *et al.* Cellular proliferation and transformation induced by HOXB4 and HOXB3 proteins involves cooperation with PBX1. *Oncogene* **16**, 3403–3412 (1998).
23. Brady, G., Barbara, M. & Iscove, N. N. Representative in vitro cDNA amplification from individual hemopoietic cells and colonies. *Mol. Cell. Biol.* **2**, 17–25 (1990).
24. Park, I.-K. *et al.* Bmi-1 is required for maintenance of adult self-renewing haematopoietic cells. *Nature* **423**, 302–305 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank S. Niessen and M. F. Clarke and colleagues for critically reading the manuscript, J. Hébert and P. Lussier for specimens and cytogenetics, J. Kros, S. Girard and N. Mayotte for technical assistance, M.-E. Leroux and S. Matte for their expertise and help regarding the maintenance and manipulation of the animals kept at the SPF facility, N. Tessier and E. Massicotte for studies with the flow cytometer, C. Charbonneau for the imaging service of IRCM, and M. van Lohuizen and R. G. Hawley for the *Bmi-1*^{-/-} mice and the MSCV vector, respectively. This work was supported by a grant from the Canadian Institute of Health Research (CIHR). J.L. was a recipient of a studentship from the CIHR and G.S. is a Scholar from the Leukemia and Lymphoma Society of America.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.S. (sauvagg@ircm.qc.ca).

An abundant population of small irregular satellites around Jupiter

Scott S. Sheppard & David C. Jewitt

Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

Irregular satellites have eccentric orbits that can be highly inclined or even retrograde relative to the equatorial planes of their planets. These objects cannot have formed by circumplanetary accretion, unlike the regular satellites that follow uninclined, nearly circular and prograde orbits¹. Rather, they are probably products of early capture from heliocentric orbits^{2–5}. Although the capture mechanism remains uncertain, the study of irregular satellites provides a window on processes operating in the young Solar System. Families of irregular satellites recently have been discovered around Saturn (thirteen members, refs 6, 7), Uranus (six, ref. 8) and Neptune (three, ref. 9). Because Jupiter is closer than the other giant planets, searches for smaller and fainter irregular satellites can be made. Here we report the discovery of 23 new irregular satellites of Jupiter, so increasing the total known population to 32. There are five distinct satellite groups, each dominated by one relatively large body. The groups were most probably produced by collisional shattering of precursor objects after capture by Jupiter.

Starting with JVI Himalia in 1904, most twentieth-century discoveries of planetary satellites were made using photographic plates. The recent development of more sensitive, large-scale charge-coupled device (CCD) detectors has refreshed the subject by enabling a new wave of satellite discovery. We have begun a systematic survey designed to assess the properties of the Jupiter satellite population. Bound satellites are confined to the region of gravitational influence of Jupiter, known as the Hill sphere. The Hill sphere radius is

$$r_H = a_J \left[\frac{m_J}{3M_\odot} \right]^{1/3} \quad (1)$$

where a_J and m_J are the semi-major axis and mass of Jupiter and $M_\odot$ is the mass of the Sun. With $a_J \approx 5 \text{ AU}$ and $m_J/M_\odot \approx 10^{-3}$, Jupiter's Hill radius is $r_H \approx 0.35 \text{ AU}$ (~ 740 Jupiter radii) corresponding to a circle of radius 4.7 degrees in the plane of the sky, when viewed at opposition (area ≈ 70 square degrees). Jupiter's Hill sphere, as seen in the plane of the sky, is 2.5 times the area of Saturn's and ten times the area of Uranus's and Neptune's. We surveyed the region around Jupiter from about 0.15 to 4.5 degrees ($0.03 \leq r/r_H \leq 0.95$), eventually finding 23 new satellites^{10,11} (Fig. 1 and Table 1). Follow-up astrometry over periods of months was used to confirm association with Jupiter, making use of orbit-fitting calculations by B. Marsden and R. Jacobson. Satellites in retrograde orbits have semi-major axes that are all less than $0.47r_H$ ($0.27r_H$ for the progrades) as is true for all known irregular satellites⁶. Some Jupiter satellites currently have apoapses up to $0.65r_H$.

At present it is practically impossible for Jupiter to capture satellites permanently because no efficient dissipation mechanism exists². Satellite capture could have occurred more easily towards the end of Jupiter's formation epoch owing to gas drag from an extended Jupiter atmosphere, the enlargement of the Hill sphere caused by the planet's mass growth and/or higher collision probabilities with nearby small bodies^{3–5}. If so, satellite capture occurred on the same (short but uncertain) timescale as Jupiter's growth, probably in the range 10^4 to 10^7 yr (refs 12, 13). The known irregular satellites are stable over the age of the Solar System, although strongly influenced by solar and planetary perturbations¹⁴.

The sizes of the new satellites can be estimated assuming geometric albedos of 4%, as measured for satellites Himalia and Elara¹⁵. We calculate radii (r) from about 1 to 4 kilometres (Table 1). Earlier observations appeared to show that the irregular satellites followed a very shallow size distribution¹⁶ with $q \approx 2$, where $n(r)dr \propto r^{-q}dr$ is the differential power-law radius distribution with $n(r)dr$ the number of satellites with radii in the range r to $r + dr$. Our observations suggest that the distribution is not well described by a single power law. We confirm $q \approx 2$ for the larger satellites ($r > 10 \text{ km}$) but find a steeper slope ($q \geq 3.5$) for those with $r < 4 \text{ km}$. There is a strong flattening in the size distribution between radii of 10 and 4 km (apparent red magnitude $19 \leq m_R \leq 21.5$). For comparison many asteroid families have $q \geq 4$ (see ref. 17), collisional equilibrium gives $q \approx 3.5$ (see ref. 18), while nonfamily asteroids show $q \approx 2.0$ to 2.5 (see ref. 19). A few large satellites are accompanied by many smaller ones

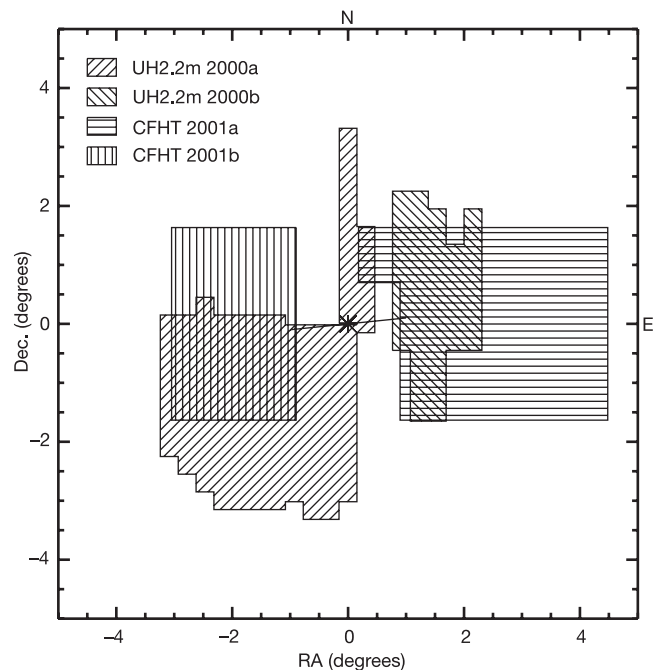


Figure 1 The area searched around Jupiter for satellites. Satellites were identified by their characteristic, Jupiter-like proper motions. We used the University of Hawaii 2.2-m telescope (UH2.2m) with its 8k CCD camera ($18.4' \times 18.4' \approx 0.094$ square degrees field-of-view) during late November 2000 and found ten new satellites (S/2000 J1 to S/2000 J10) (see ref. 10). In mid-December 2001 we used the Canada–France–Hawaii 3.6-m telescope (CFHT) with the facility 12k CCD camera ($43' \times 28' \approx 0.33$ square degrees) and discovered eleven new satellites (S/2001 J1 to S/2001 J11) (see ref. 11). An additional two satellites were discovered in the survey fields during follow-up observations, for a total of 23 new satellites. The thick black line shows the orientation of the ecliptic during the observations. All the search fields of 2000 were examined both by visual blinking and a computer algorithm written by E. Magnier. For the 2001 fields we used a computer program written by J. Kleyna along with visual blinking to find the satellites. A measure of completeness of our survey is provided by our serendipitous detections of previously known satellites. We detected all six of the then-known nine irregular satellites of Jupiter that were in our fields in November 2000. In December 2001 we detected 12 of the 15 irregulars predicted to lie within the survey area, with the other three undetected because they were near bright stars. Because weather factors varied over the different nights we split the survey up into four parts. UH2.2m 2000a: Total area searched was 12.0 square degrees to a limiting R -magnitude of about 21.5 with one new satellite discovered. UH2.2m 2000b: 4.4 square degrees, $m_R \approx 22.5$, nine new satellites. CFHT 2001a: 12.4 square degrees, $m_R \approx 23.2$, ten new satellites. CFHT 2001b: 6.7 square degrees, $m_R \approx 22.5$, one new satellite. Both S/2000 J11 and S/2002 J1 were discovered while recovering other satellites and thus are not included in the above survey statistics.

($r < 4$ km), the latter having a size distribution similar to that of the main-belt asteroid families. By extrapolation, we estimate that, within a factor of two, Jupiter possesses approximately 100 irregular satellites with $r \geq 0.5$ km, corresponding to about 24th magnitude. We also find that the completeness limit for satellites at Jupiter is currently $r \approx 3$ km ($m_R \approx 21.5$).

The mean orbital inclination and mean semi-major axis distributions show two distinct prograde groups, here named the Himalia and Themisto groups after the largest object in each (Fig. 2). Clumping in Fig. 2 suggests that the retrograde satellites comprise at least three groups (we denote these the Pasiphae, Ananke and Carme groups). The absence of objects with intermediate inclinations $55^\circ < i < 130^\circ$ may be a result of the Kozai effect¹⁴, which couples variations in inclination and eccentricity and forces highly inclined satellites to dip into the galilean satellite region, where they are removed by collisions.

These dynamical groups suggest origin by the break-up of multiple parent objects. Each retrograde group contains one large object ($r > 14$ km) along with several smaller ones ($r < 4$ km). Sinope, the only other large satellite, may currently be the only known member of a fourth group or it may be part of the Pasiphae group. In support of a collisional origin, we note that the dispersion velocities²⁰ between members of each group are comparable to the escape velocity of the largest body (~ 30 m s⁻¹ for each retrograde group). Velocity differences between objects in different groups (> 200 m s⁻¹), probably from different parents, are much larger than the escape velocity from any member. Also, observations of the brightest irregulars show that the retrogrades are redder and more spectrally diverse than the Himalia prograde group²¹, consistent

with the origin of the former from the break-up of four different parent objects.

The sizes of the parent bodies can be estimated by combining the volumes of the satellites within each group. We infer that the Ananke group parent body had a radius of about $r_p \approx 14$ km while the ratio of the mass in the largest fragment to the mass of the parent is $M_1/M_p \approx 0.98$. For the Carme group we find $r_p \approx 23$ km with $M_1/M_p \approx 0.99$ and for the Pasiphae group $r_p \approx 30$ with $M_1/M_p \approx 0.99$ (or 0.79 if including Sinope). For the prograde Himalia group we find $r_p \approx 89$ km with $M_1/M_p \approx 0.87$. Satellite groupings with $M_1/M_p \approx 1$ suggest collisional origin by shattering with a projectile barely above the disruption threshold. Such high mass ratios are atypical of asteroid families¹⁷, although the Vesta asteroid family, with $M_1/M_p \approx 0.95$, may be similar²² in suffering only minimal damage during its lifetime from relatively small impacts.

It is unlikely that the satellite groups were produced by the sole action of aerodynamic forces during capture, because self-gravity would prevent the fragments from dispersing⁴. Additionally, gas drag acting on the fragments would produce a size-dependent sorting of the orbits within each group^{2,4,6}. No size versus orbital property correlations are seen in the groupings in the ~ 1 km to ~ 100 km size range. For these reasons we believe that the disruptions occurred after capture and after the dissipation of the gas left over from Jupiter's formation.

Fragmentation of the parent satellites could be caused by impact with interplanetary projectiles (principally comets) or by collision with other satellites. Break-up requires that the projectile kinetic energy should exceed the gravitational binding energy of the parent.

Table 1 Physical and orbital properties of the irregular satellites*

Name	Semi-major axis, a (km)	Inclination, i (degrees)	Eccentricity, e	Period (days)	Mag. (m_R)	Diameter (km)	Year of discovery
Themisto Group	Prograde						
XVIII Themisto	7,507,000	43.08	0.24	130.0	21.0	8	2000
Himalia Group	Prograde						
XIII Leda	11,165,000	27.46	0.16	240.9	19.2	20	1974
VI Himalia	11,461,000	27.50	0.16	250.6	14.2	170	1904
X Lysithea	11,717,000	28.30	0.11	259.2	17.9	36	1938
VII Elara	11,741,000	26.63	0.22	259.6	16.0	86	1905
S/2000 J11	12,555,000	28.30	0.25	287.0	22.4	4	2000
Ananke Group	Retrograde						
S/2001 J10	19,302,000	145.8	0.14	550.7	23.1	2	2001
S/2001 J7	21,027,000	148.9	0.23	620.0	22.8	3	2001
XXII Harpalyke	21,105,000	148.6	0.23	623.3	22.2	4	2000
XXVII Praxidike	21,147,000	149.0	0.23	625.3	21.2	7	2000
S/2001 J9	21,168,000	146.0	0.28	623.0	23.1	2	2001
S/2001 J3	21,252,000	150.7	0.21	631.9	22.1	4	2001
XXIV Iocaste	21,269,000	149.4	0.22	631.5	21.8	5	2000
XII Ananke	21,276,000	148.9	0.24	610.5	18.3	28	1951
S/2001 J2	21,312,000	148.5	0.23	632.4	22.3	4	2001
Carme Group	Retrograde						
S/2001 J6	23,029,000	165.1	0.27	716.3	23.2	2	2001
S/2002 J1	23,064,000	165.0	0.26	715.6	22.8	3	2002
S/2001 J8	23,124,000	165.0	0.27	720.9	23.0	2	2001
XXI Chaldene	23,179,000	165.2	0.25	723.8	22.5	4	2000
XXVI Isonoe	23,217,000	165.2	0.25	725.5	22.5	4	2000
XXV Erinome	23,279,000	164.9	0.27	728.3	22.8	3	2000
XX Taygete	23,360,000	165.2	0.25	732.2	21.9	5	2000
XI Carme	23,404,000	164.9	0.25	702.3	17.1	46	1938
S/2001 J11	23,547,000	165.2	0.26	741.0	22.7	3	2001
XXIII Kalyke	23,583,000	165.2	0.25	743.0	21.8	5	2000
Pasiphae Group	Retrograde						
S/2001 J4	23,219,000	150.4	0.28	720.8	22.7	3	2001
VIII Pasiphae	23,624,000	151.4	0.41	708.0	16.6	60	1908
XIX Megaclyte	23,806,000	152.8	0.42	752.8	21.7	6	2000
S/2001 J5	23,808,000	151.0	0.31	749.1	23.0	2	2001
IX Sinope	23,939,000	158.1	0.25	724.5	17.6	38	1914
XVII Callirhoe	24,102,000	147.1	0.28	758.8	20.3	8	1999
S/2001 J1	24,122,000	152.4	0.32	765.1	22.0	4	2001

* Orbital data are from R. Jacobson (http://ssd.jpl.nasa.gov/sat_elem.html); fits are over a 1,000-year time span. a , mean semi-major axis with respect to Jupiter; i , mean inclination of orbit with respect to Jupiter's equator; e , mean eccentricity; period, orbital period of satellite around Jupiter; mag., apparent red (0.65 μ m wavelength) magnitude; diameter, diameter of satellite, assuming a geometric albedo of 0.04.

If half of the kinetic energy of the comet goes into breaking apart the target satellite²³ and if the collision velocity is about 5 km s^{-1} , a target satellite with $r = 25 \text{ km}$ must be struck by a $r \approx 1 \text{ km}$ projectile in order to be disrupted. At the present epoch, the flux of Jupiter-crossing comets of this size is 10^3 to 10^4 times too small to shatter the irregular satellite parent bodies²⁴. However, lunar crater counts show a very rapid fall in the projectile flux in the first few 100 Myr, approaching the current steady state flux about 3.5×10^9 years ago²⁵. The time-averaged flux of lunar impactors in the last 4.5×10^9 years is roughly 10^5 times the current flux. If applicable at Jupiter, this large initial flux would be sufficient to shatter each of the parent satellites and could explain the observed groups. 'Particle-in-a-box' type calculations show that the time-scales for collision among the known satellites are very long. For the retrograde satellites the timescale ($\sim 10^{11}$ years) is much larger than the age of the Solar System, while the progrades have a shorter, but still long, collision timescale ($\sim 10^9$ years). Thus, fragmentation by collision among the current satellites seems unlikely. However, it is possible that Jupiter once held a much larger population of irregular satellites and that the observed groups are merely the products of disruptive, post-capture collisions amongst them.

Although Jupiter currently possesses the largest number of known satellites of any of the planets, the other giant planets also have significant irregular satellite populations. All are thought to have been captured by their respective planets^{6,8}. In common with Jupiter, the other systems show dynamical grouping, but the satellites of more distant planets are fainter and observationally less well characterized. If Jupiter were displaced to the distances of the other giant planets we would only be able to detect about 11 of the 32 known irregular satellites at Saturn's distance, six at Uranus's

distance, and two at Neptune's distance in a survey of each Hill sphere to 24th magnitude (corresponding to radii of 3.5, 18, and 40 km, respectively). In fact, Saturn has 13 known irregulars within three or four distinct dynamical groups, Uranus has five with one or two distinct groups, and Neptune has two irregulars brighter than 24th magnitude. By this measure, the four giant planets possess about the same number of irregular satellites and satellite groups with no dependence on the planet's mass. This is especially remarkable given that the ice giants Uranus and Neptune may have had formation histories quite different from the gas giants Jupiter and Saturn^{13,26}.

Note added in proof: A number of new small irregular satellites were discovered after the submission of this paper (S/2003 J1 to J20; refs 7, 27). With the exception of S/2003 J20, the new satellites fall into the retrograde dynamical groupings described in this paper. The orbital elements of the new objects are still uncertain and therefore have not been fully incorporated into the paper. S/2003 J20 has a prograde orbit unlike any other known satellite ($a \approx 17,100,000 \text{ km}$ ($0.33 R_H$), $i \approx 55^\circ$, $r \approx 1.5 \text{ km}$). A few of the new discoveries may have inclinations and semi-major axes similar to Sinope. □

Received 2 December 2002; accepted 14 March 2003; doi:10.1038/nature01584.

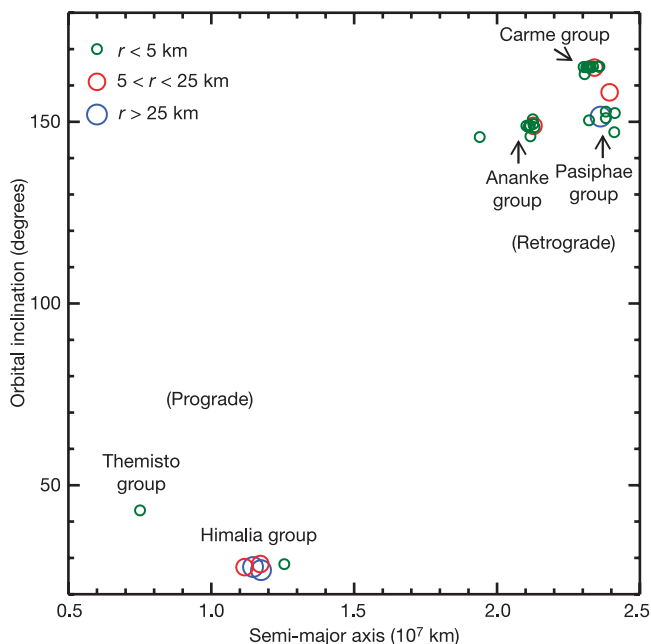


Figure 2 The mean semi-major axes versus the mean inclinations of the irregular satellites of Jupiter. Five distinct clusters are evident with the Ananke, Pasiphae and Carme retrograde groups at about 300 Jupiter radii with mean inclinations around 149, 151 and 165 degrees respectively, the Himalia (main) prograde group at about 155 Jupiter radii with inclinations near 28 degrees, and the Themisto prograde group with Themisto as the lone member at about 100 Jupiter radii and inclination of 45 degrees. Satellite symbols are plotted according to size bins. S/2001 J10 is the retrograde satellite to the left of the Ananke group. Its orbit, along with Sinope's and Pasiphae's, may have been modified through resonances (R. Jacobson, personal communication; ref. 28). The satellites discovered in 2000 and 2001 have been observed for at least two oppositions, except for S/2000 J11 which has only about a month of observations.

1. Peale, S. Origin and evolution of the natural satellites. *Annu. Rev. Astron. Astrophys.* **37**, 533–602 (1999).
2. Kuiper, G. On the origin of the satellites and the Trojans. *Vistas Astron.* **2**, 1631–1666 (1956).
3. Colombo, G. & Franklin, F. On the formation of the outer satellite groups of Jupiter. *Icarus* **15**, 186–189 (1971).
4. Pollack, J., Burns, J. & Tauber, M. Gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* **37**, 587–611 (1979).
5. Heppenheimer, T. & Porco, C. New contributions to the problem of capture. *Icarus* **30**, 385–401 (1977).
6. Gladman, B. et al. Discovery of 12 satellites of Saturn exhibiting orbital clustering. *Nature* **412**, 163–166 (2001).
7. Sheppard, S. et al. Satellites of Jupiter and Saturn. *IAU Circ. No.* 8116 (2003).
8. Gladman, B. et al. The discovery of Uranus XIX, XX and XXI. *Icarus* **147**, 320–324 (2000).
9. Holman, M. et al. Satellites of Neptune. *IAU Circ. No.* 8047 (2003).
10. Sheppard, S., Jewitt, D., Fernandez, Y., Magnier, E. & Marsden, B. Satellites of Jupiter. *IAU Circ. No.* 7555 (2001).
11. Sheppard, S., Jewitt, D., Kleyna, J., Marsden, B. & Jacobson, R. Satellites of Jupiter. *IAU Circ. No.* 7900 (2002).
12. Pollack, J. et al. Formation of the giant planets by concurrent accretion of solids and gas. *Icarus* **124**, 62–85 (1996).
13. Boss, A., Wetherill, G. & Haghighipour, N. Rapid formation of ice giant planets. *Icarus* **156**, 291–295 (2002).
14. Carruba, V., Burns, J., Nicholson, P. & Gladman, B. On the inclination distribution of the Jovian irregular satellites. *Icarus* **158**, 434–449 (2002).
15. Cruikshank, D. Radii and albedos of four trojan asteroids and Jovian satellites 6 and 7. *Icarus* **30**, 224–230 (1977).
16. Gehrels, T. In *Comets, Asteroids, Meteorites* (ed. Delsemme, A.) 323–326 (Univ. Toledo Press, Toledo, 1977).
17. Tanga, P. et al. On the size distribution of asteroid families: The role of geometry. *Icarus* **141**, 65–78 (1999).
18. Dohnanyi, J. Collisional models of asteroids and their debris. *J. Geophys. Res.* **74**, 2531–2554 (1969).
19. Cellino, A., Zappala, V. & Farinella, P. The size distribution of mainbelt from IRAS data. *Mon. Not. R. Astron. Soc.* **253**, 561–574 (1991).
20. Kessler, D. Derivation of the collision probability between orbiting objects: The lifetimes of Jupiter's outer moons. *Icarus* **48**, 39–48 (1981).
21. Rettig, T., Walsh, K. & Consolmagno, G. Implied evolutionary differences of the Jovian irregular satellites from a BVR colour survey. *Icarus* **154**, 313–320 (2001).
22. Thomas, P. et al. Impact excavation on asteroid 4 Vesta: Hubble Space Telescope results. *Science* **277**, 1492–1495 (1997).
23. Melosh, H. *Impact Cratering a Geologic Process* 46–86 (Oxford Univ. Press, Oxford, 1989).
24. Nakamura, T. & Yoshikawa, M. Close encounters and collisions of short-period comets with Jupiter and its satellites. *Icarus* **116**, 113–130 (1995).
25. Hartmann, W., Ryder, G., Dones, L. & Grinspoon, D. In *Origin of the Earth and Moon* (eds Canup, R. & Righter, K.) 493–512 (Univ. Arizona Press, Tucson, 2000).
26. Thommes, E., Duncan, M. & Levison, H. The formation of Uranus and Neptune among Jupiter and Saturn. *Astron. J.* **123**, 2862–2883 (2002).
27. Sheppard, S. et al. Satellites of Jupiter. *IAU Circ. No.* 8089 (2003).
28. Saha, P. & Tremaine, S. The orbits of the retrograde Jovian satellites. *Icarus* **106**, 549–562 (1993).

Acknowledgements We thank Y. Fernandez for help with the observations. The Canada–France–Hawaii telescope is operated by the National Research Council of Canada, Le Centre National de la Recherche Scientifique de France, and the University of Hawaii. This work was supported by a grant to D.C.J. from NASA.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.S.S. (sheppard@ifa.hawaii.edu).

Chaos-assisted capture of irregular moons

Sergey A. Astakhov*, Andrew D. Burbanks†, Stephen Wiggins† & David Farrelly*

* Department of Chemistry & Biochemistry, Utah State University, Logan, Utah 84322-0300, USA

† School of Mathematics, University of Bristol, Bristol BS8 1TW, UK

It has been thought^{1–3} that the capture of irregular moons—without circular orbits—by giant planets occurs by a process in which they are first temporarily trapped by gravity inside the planet's Hill sphere (the region where planetary gravity dominates over solar tides⁴). The capture of the moons is then made permanent by dissipative energy loss (for example, gas drag³) or planetary growth². But the observed distributions of orbital inclinations, which now include numerous newly discovered moons^{5–8}, cannot be explained using current models. Here we show that irregular satellites are captured in a thin spatial region where orbits are chaotic⁹, and that the resulting orbit is either prograde or retrograde depending on the initial energy. Dissipation then switches these long-lived chaotic orbits¹⁰ into nearby regular (non-chaotic) zones from which escape is impossible. The chaotic layer therefore dictates the final inclinations of the captured moons. We confirm this with three-dimensional Monte Carlo simulations that include nebular drag^{3,4,11}, and find good agreement with the observed inclination distributions of irregular moons at Jupiter⁷ and Saturn⁸. In particular, Saturn has more prograde irregular moons than Jupiter, which we can explain as a result of the chaotic prograde progenitors being more efficiently swept away from Jupiter by its galilean moons.

The recent discoveries of large numbers of irregular satellites at Jupiter^{5–7} and Saturn⁸ are helping to shed new light on how planets capture other bodies^{1–3,12–22}. It has been widely held, for example, that the observed preponderance of retrograde irregular moons at Jupiter is due mainly to the well known enhanced stability of retrograde orbits with large semimajor axis, a (refs 2, 20–22). However, Saturn's family of irregulars is now known to contain a much more even mix of prograde and retrograde satellites than that of Jupiter^{5–8}. Further, Saturn's irregulars have similarly large semimajor axes (as compared to Jupiter's moons) when expressed in planetary radii⁸. This rules out simple stability arguments as explanations for the markedly different distributions of moons observed at the two planets.

Here we study capture in the three-dimensional (3D) 'spatial' circular restricted three-body problem (CRTBP), taking the Sun–Jupiter–moon system as our primary example⁴. In a coordinate system rotating with the mean motion, but with the origin transformed to the planet, the spatial CRTBP hamiltonian⁴ becomes:

$$H = E = \frac{1}{2} \mathbf{p}^2 - (xp_y - yp_x) - \frac{\mu}{\sqrt{x^2 + y^2 + z^2}} - \frac{1 - \mu}{\sqrt{(1+x)^2 + y^2 + z^2}} - (1 - \mu)x + \alpha \quad (1)$$

where $\mu = m_1/(m_1 + m_2)$, and m_1 and m_2 are the masses of the primaries; α is a collection of inessential constants; the total energy in the non-inertial frame, E , is related to the Jacobi constant through $C_1 = -2E$; and $\mathbf{r} = (x, y, z)$ and $\mathbf{p} = (p_x, p_y, p_z)$ are the coordinate and momentum vectors of the test particle. Angular momentum, $\mathbf{h} = (h_x, h_y, h_z)$ where $h_z = xp_y - yp_x$, is now defined with respect to the planet, as is most natural in a study of capture²³.

Figure 1a and b shows surfaces of zero velocity⁴ at two energies for the 3D Sun–Jupiter system, together with the Hill sphere. Surfaces of zero velocity limit the motion in the rotating frame, and so serve to define an energetically accessible 'bubble' that may intersect the Hill sphere⁴. At both energies in Fig. 1, the two Lagrange saddle points, L_1 and L_2 , are 'open', and act as gateways between the interior energy bubble and heliocentric orbits²⁴. Figure 1c and d shows orbital inclination distributions⁴, $i = \cos^{-1}(h_z/|\mathbf{h}|)$, of test particles as they pass through the Hill sphere at the same two energies (see Methods). The key finding is that at energies slightly above the Lagrange points, only prograde orbits can enter (or leave) the capture zone. At higher capture energies, the distribution shifts to include both senses of h_z . The statistics of inclination distributions will, therefore, be expected to depend strongly on energy, that is, the geometry of how (and if) the curves of zero velocity intersect the Hill sphere.

Figure 2 portrays the structure of phase space in the planar limit ($z = p_z = 0$) in a series of Poincaré surfaces of section⁹ (SOS) at four energies. At the lowest energy (Fig. 2a), many of the prograde orbits are chaotic, whereas all the retrograde orbits are regular⁹. Because incoming orbits cannot penetrate the regular Kolmogorov–Arnold–Moser (KAM) regions⁹, prograde orbits must remain prograde. Although KAM tori in 3D cannot 'block' trajectories in this manner, orbits can only enter nearly integrable volumes of phase space by Arnold diffusion, which, by the Nekhoroshev theorem, is expected to occur exponentially slowly^{9,10}. In essence, the chaotic layer selects for the sense of the angular momentum of incoming and outgoing particles.

After L_2 has opened up (Fig. 2b), the chaotic 'sea' of prograde orbits quickly 'evaporates', except for a thin, advancing front of chaos which clings to the KAM tori, buffering them from the expanding basin of direct scattering. With increasing energy, this

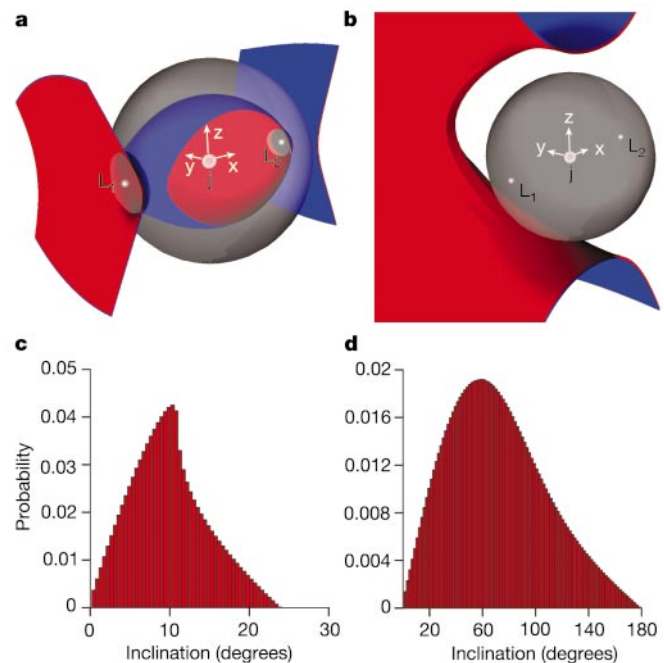


Figure 1 The intersection of the Hill sphere in 3D with the surfaces of zero velocity, and histograms of inclination distributions for 10^8 test particles originating at the Hill sphere. Data are shown for two energies (**a** and **c**, **b** and **d**) for the Sun–Jupiter system with $\mu = 9.5358 \times 10^{-4}$. In **a** and **c** the Jacobi constant $C_J = 3.0380$, and in **b** and **d** $C_J = 3.0100$. The Lagrange saddle points are L_1 and L_2 . The coordinate system is centred on Jupiter (small pink sphere, not to scale, labelled 'j'). The Hill sphere is the large grey sphere.

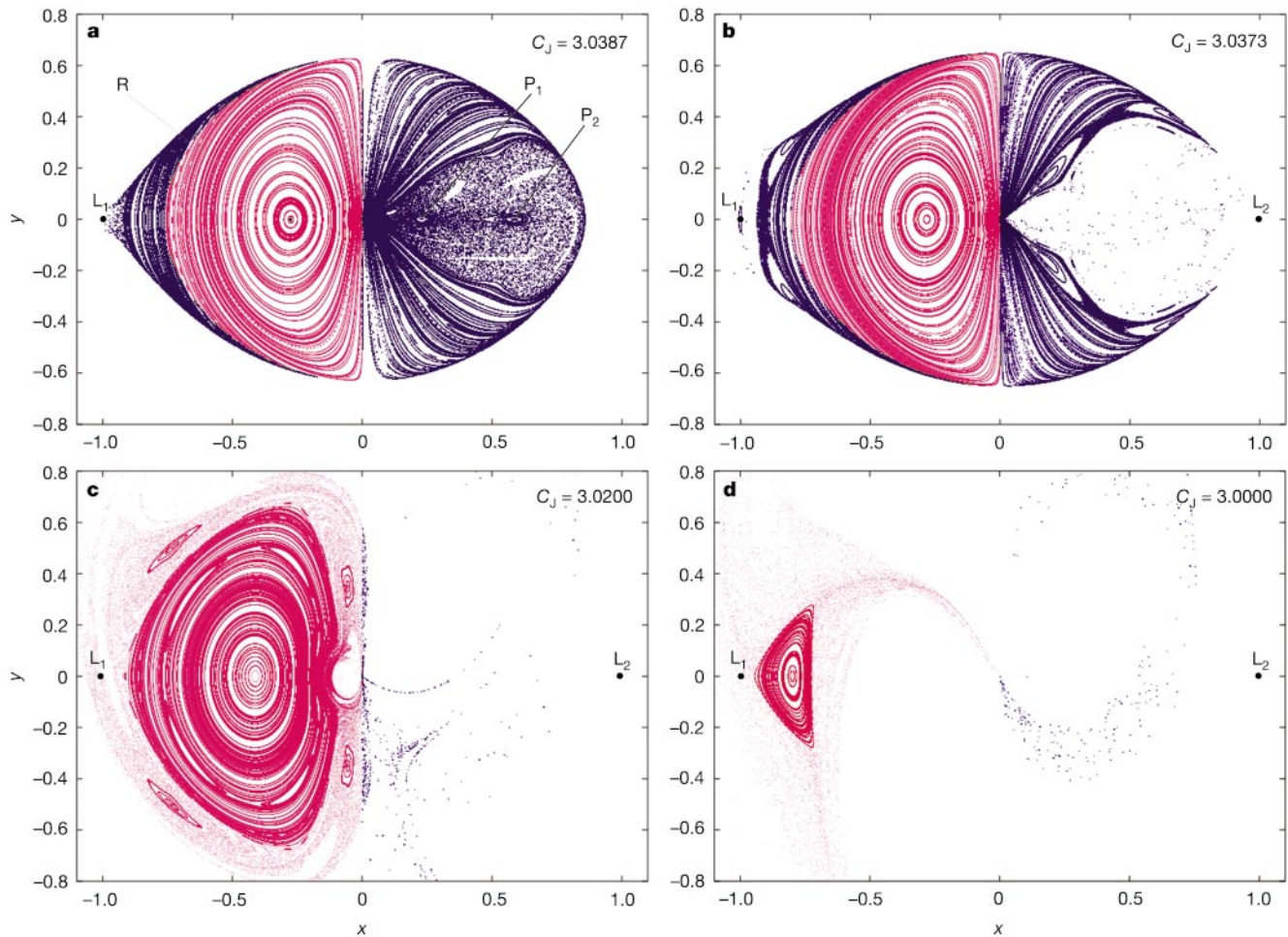


Figure 2 Poincaré surfaces of section showing regions of chaotic ('shotgun') pattern and regular (nested curves) dynamics. The surfaces are shown for randomly chosen initial conditions, and for the four indicated values of C_J ; units have been rescaled to $R_H = 1$ (ref. 4). Points on the surface are coloured according to the sign of their angular

momentum h_z as they penetrate the x - y plane; purple, prograde ($h_z > 0$); red, retrograde ($h_z < 0$). In **a**, P_1 , P_2 and R indicate two prograde and one retrograde periodic orbit, respectively.

front smoothly shifts from prograde to retrograde motion. The surviving frontier tori are 'sticky', and chaotic orbits near them can become trapped in almost regular orbits for very long times¹⁰. Note that the KAM tori in Fig. 2 exist at energies above L_1 and L_2 , and so dissipation need only be sufficient to switch chaotic orbits into the nearest KAM region for permanent capture to happen. At high energy, permanent capture is almost exclusively into retrograde KAM tori surrounding the almost circular, retrograde orbit, labelled R in Fig. 2a (refs 20, 21). The relative stability of prograde and retrograde orbits in two dimensions (2D) can be understood using Kaptza averaging (see Methods).

Figure 3 shows the normalized probability distribution of orbits from 3D Monte Carlo simulations which survive a cut-off time of 20,000 years; the distribution is shown as a function of the Jacobi constant and final inclination of the orbits. A clear trend from prograde to retrograde trapping with increasing total energy is found, which is consistent with Fig. 2. The three main structures visible in Fig. 3 correspond to low (prograde), intermediate and high (retrograde) inclinations. The large island at $i \approx 100$ – 120° is related to librational motion around the argument of perijove $\omega = \pm 90^\circ$, that is, the Kozai resonance^{16,24}. We confirmed numerically that these distributions are similar for all the giant planets.

To understand further the role of chaos in shaping capture, we consider Hill's approximation, which is valid (as here) for $\mu \ll 1$; in appropriately scaled units, Hill's equations contain no parameters, and so test-particle trajectories will scale accordingly for all the giant

planets⁴. This is confirmed by a comparison of SOS for the giant planets. Thus, the observed differences between Jupiter's and Saturn's families of irregulars probably cannot be explained using hamiltonian dynamics alone. However, significant differences may arise when a

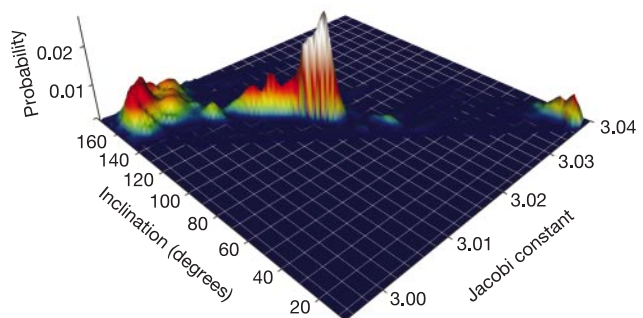


Figure 3 The normalized probability distribution of 3D orbits that survived in Monte Carlo simulations (without dissipation) for 20,000 years. Data are shown as a function of their final inclination and initial value of the Jacobi constant. 80 million test particles were integrated, of which $\sim 38,000$ survived for 5,000 years and $\sim 23,000$ survived for 20,000 years. Using time-averaged inclinations with moving windows of 100–1,000 years produces essentially identical distributions. In these simulations, test particles were removed if they came within two planetary radii of the planet.

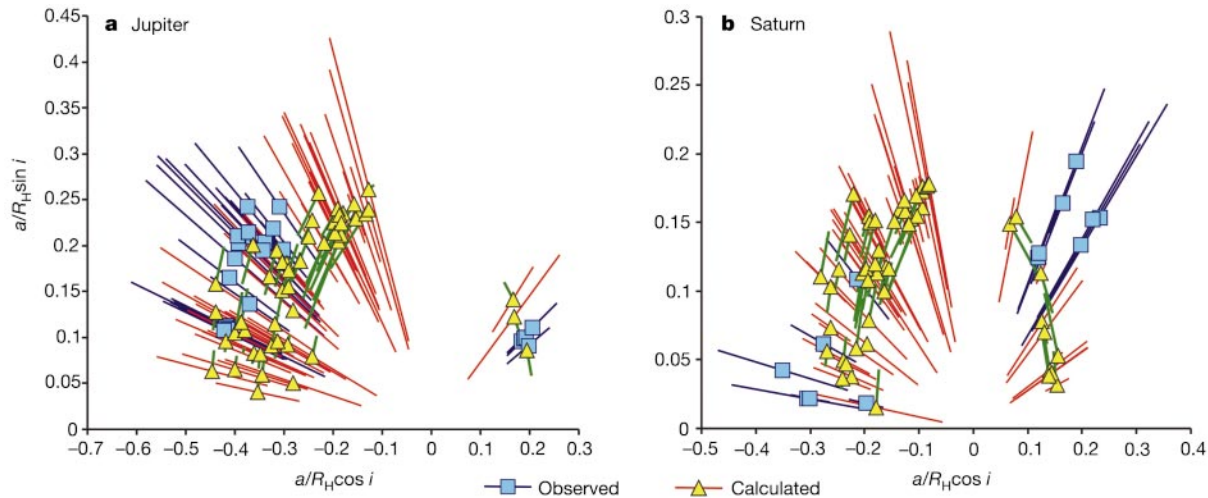


Figure 4 Comparison of the orbital properties of the known irregular satellites of Jupiter^{5,6} (to 2002⁷) and Saturn⁸ with dissipative Monte Carlo simulations. The radial distance from the origin represents the moon's orbital semimajor axis, a , as a fraction of R_H . The length of each line shows the pericentre-to-apocentre distance of the orbit. Time-averaged

orbital elements are used. The green lines show the total excursion in inclination undergone by each moon. In most cases, a strong correlation between initial and final inclination is observed. In some cases, initial and final inclinations are indistinguishable on this scale.

particle's motion is dependent on additional perturbations that do not follow Hill's scaling (for example, nebular gas-drag^{4,11}).

In addition, the local environment at each planet is different; one such variable is the volume of the Hill sphere occupied by the massive, regular moons. This observation is important, because prograde orbits penetrate much deeper towards the planet than do most of the retrograde orbits (Fig. 2b and d). Therefore, to be permanently captured, prograde orbits (especially) must survive close encounters or collisions with the massive regular moons. Orbits in the chaotic layer may be very long lived, and so can cross the orbits of the inner satellites many times; and because chaotic orbits are rather sensitive to perturbations⁹, such passages will strongly perturb their orbital elements. Therefore, in our simulations with dissipation, we eliminated test particles that crossed the outermost massive regular moons of Jupiter and Saturn. We decided to use Saturn's moon Titan (which is similar in mass to Jupiter's outermost moon Callisto), rather than Saturn's actual outermost, but significantly smaller, regular moon Iapetus⁴.

The results of these simulations (see Methods) are shown in Fig. 4. Saturn has a clear tendency to capture a higher ratio of prograde to retrograde moons as compared to Jupiter. This can be traced to the fact that Callisto's semimajor axis is a larger fraction of the Hill radius for Jupiter (3.55%) than is the case for Titan, whose orbit is only 1.84% of Saturn's Hill radius. This effect was confirmed in the simulations by moving Titan's orbit out to the same relative distance (as a fraction of the Hill radius) as Callisto, which resulted in a prograde-poor distribution at Saturn, similar to that at Jupiter. Thus, the relative scarcity of jovian prograde irregulars may be due to their chaotic prograde progenitors having been swept away more efficiently by Jupiter's galilean moons.

The inclinations of captured test particles (Fig. 4) compare well with observed inclinations, and similarly seem to occur in definite clusters and sub-clusters. All of these clusters originated in initially chaotic orbits, which originated on the Hill sphere. Remarkably, these orbits nevertheless tend to preserve their initial inclinations. This phenomenon of 'inclination memory' suggests that the observed clusters of irregular moons might be relics of the capture dynamics, rather than being necessarily the result of post-capture fragmentation through collisions^{3,8,12}. For example, whereas average satellite inclinations are clustered, other orbital elements such as average eccentricity may be more widely dispersed; this is more

difficult to explain in a collision model⁵⁻⁸. A test of the fragmentation hypothesis will come from more detailed observations of the physical characteristics of irregular moons. These observations will also be a test of our model, which provides a single, unified, mechanism for the capture of both prograde and retrograde moons^{2,15,20-22}.

Chaos-assisted capture may be involved in the formation of asteroid or Kuiper-belt binaries²⁵, and also has implications for the feasibility of the capture of massive bodies such as Neptune's retrograde moon Triton²⁶. Figure 3 indicates that a moon captured into a high-inclination orbit would probably have first been trapped in the chaotic layer close to an almost-circular orbit (labelled R in Fig. 2); this provides a possible clue to the origin of Triton's circular, highly inclined orbit²⁶. □

Methods

Monte Carlo simulations

An isotropic flux of test particles was originated in 3D on the Hill sphere (radius $R_H = a_p(\mu/3)^{1/3}$ where a_p is the planet's semimajor axis), and integrated until one of the following occurred: the particle left the Hill sphere; it penetrated a sphere, centred on the planet, of a given radius (see figure captions and text); or it survived for a predetermined cut-off time. Initial conditions were generated as follows: the vector $\mathbf{r}$ was chosen uniformly and randomly on the surface of the Hill sphere. Velocities were also chosen uniformly and randomly in accordance with the value of the Jacobi constant. This constant was chosen randomly and uniformly: $C_J \in (2.995, C_J^{(1)})$, where $C_J^{(1)}$ is the value of C_J at L_1 . Figure 1c and d show the resulting distributions of initial inclinations for these initial conditions. Numerical integrations were done using a Bulirsch-Stoer procedure²⁷. Choosing particles on the Hill sphere, where the motion is chaotic or scattering²², minimizes the risk of accidentally starting orbits deep inside impenetrable KAM regions¹⁸. Although permanently bound, such orbits could never in fact have been trapped, because KAM regions cannot be penetrated in 2D and only exponentially slowly in 3D. It is only in the thin chaotic layer between scattering and stability that capture can happen. This method of selecting initial conditions thus specifically selects for these particles.

Surfaces of section

The SOS⁹ were computed by integrating randomly chosen ensembles of test particles with initial conditions chosen inside the Hill radius (R_H) and integrated using Levi-Civita regularizing coordinates²⁸. The SOS chosen is the x - y plane with $p_x = 0$ and $\dot{y} > 0$; points have been coloured in Fig. 2 according to the sign of their angular momentum as they intersect that surface. In almost all cases the sign of h_z is well conserved.

Simulations with dissipation

We integrated trajectories including a dissipative force representing nebular drag⁴, $\mathbf{F}_{\text{drag}} = -\gamma\mathbf{v} = -\gamma(\dot{x}, \dot{y}, \dot{z})$, where γ is the drag parameter. This form was chosen not only for its simplicity, but also because it scales in the Hill approximation, which allowed for a more even-handed comparison of Jupiter and Saturn. This is true even in the CRTBP, for which

Hill's scaling is only approximate. Experimental simulations revealed that, for fixed integration times, the main effect of changing γ was to influence the final average values of the semimajor axes of captured moons with respect to the planet; in contrast, the final inclination distributions were quite robust to the precise value of this parameter. Thus we chose γ heuristically so that the captured moons ended up roughly in the observed semimajor axis ranges; we set $\gamma = 7 \times 10^{-5}$ in units scaled for Jupiter, and $\gamma = 2 \times 10^{-4}$ in units scaled for Saturn.

Integrations were performed for both Jupiter and Saturn for a maximum of 10,000 years for each test particle. Integrations were stopped, as explained in the text, if test particles crossed the orbit of Callisto (at Jupiter) or Titan (at Saturn), or if they left the Hill sphere. The simulations reported in Fig. 4 were stopped when 50 moons had been captured, but computations in which several thousand moons were captured produce similar results. We have also performed parallel simulations in the elliptic restricted three-body problem for Jupiter and Saturn, and also used different forms of dissipation—for example, nebular gas-drag, $\mathbf{F}_{\text{drag}} = -\gamma|\mathbf{v}|\mathbf{v}$ (ref. 11). All of these variations produced comparable results.

Kapitza averaging

The relative stability of prograde and retrograde orbits in 2D can be understood qualitatively by Taylor expansion of the solar part of the CRTBP hamiltonian, followed by Kapitza averaging in plane polar coordinates over the angle ϕ conjugate to h_z . This is similar to the analogous problem of ionization (escape) of an electron from a hydrogen atom in a rotating field^{29,30}. As in the atomic problem, this strategy produces an effective potential whose saddle point is higher for one sense of angular momentum: in this case, the retrograde orbits, which are therefore more stable than the prograde orbits. Further, using methods similar to those in ref. 23, it is possible to show that h_z is the lowest-order term in an approximate 'third-integral' valid inside the Hill sphere.

Received 6 December 2002; accepted 26 March 2003; doi:10.1038/nature01622.

1. Peale, S. J. Origin and evolution of the natural satellites. *Annu. Rev. Astron. Astrophys.* **37**, 533–602 (1999).
2. Heppenheimer, T. A. & Porco, C. New contributions to the problem of capture. *Icarus* **30**, 385–401 (1977).
3. Pollack, J. R., Burns, J. A. & Tauber, M. E. Gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* **37**, 587–611 (1979).
4. Murray, C. D. & Dermot, S. F. *Solar System Dynamics* (Cambridge Univ. Press, Cambridge, 1999).
5. Shephard, S. S., Jewitt, D. C., Fernandez, Y. R., Magnier, G. & Marsden, B. G. Satellites of Jupiter. *IAU Circ. No.*, 7555 (2001).
6. Shephard, S. S., Jewitt, D. C., Kleyna, J., Marsden, B. G. & Jacobson, R. Satellites of Jupiter. *IAU Circ. No.*, 7900 (2002).
7. Shephard, S. S. *et al.* Satellites of Jupiter. *IAU Circ. No.*, 8087 (2003).
8. Gladman, B. J. *et al.* Discovery of 12 satellites of Saturn exhibiting orbital clustering. *Nature* **412**, 163–166 (2001).
9. Lichtenberg, A. J. & Leiberman, M. A. *Regular and Chaotic Dynamics*, 2nd edn 174–183 (Springer, New York, 1992).
10. Perry, A. D. & Wiggins, S. KAM tori are very sticky: rigorous lower bounds on the time to move from an invariant Lagrangian torus with linear flow. *Physica D* **71**, 102–121 (1994).
11. Kary, D. M., Lissauer, J. J. & Greenzweig, Y. Nebular gas drag and planetary accretion. *Icarus* **106**, 288–307 (1993).
12. Colombo, G. & Franklin, F. A. On the formation of the outer satellite group of Jupiter. *Icarus* **15**, 186–189 (1971).
13. Huang, T.-Y. & Innanen, K. A. The gravitational escape/capture of planetary satellites. *Astron. J.* **88**, 1537–1547 (1983).
14. Murison, M. A. The fractal dynamics of satellite capture in the circular restricted three-body problem. *Astron. J.* **98**, 2346–2359 (1989).
15. Namouni, F. Secular interactions of coorbiting objects. *Icarus* **137**, 293–314 (1999).
16. Carruba, V., Burns, J. A., Nicholson, P. D. & Gladman, B. J. On the inclination distribution of the Jovian irregular satellites. *Icarus* **158**, 434–449 (2002).
17. Nesvorný, D., Thomas, F., Ferraz-Mello, S. & Morbidelli, A. A perturbative treatment of co-orbital motion. *Celest. Mech. Dynam. Astron.* **82**, 323–361 (2002).
18. Vieira Neto, E. & Winter, O. C. Time analysis for temporary gravitational capture: satellites of Uranus. *Astron. J.* **122**, 440–448 (2001).
19. Marzani, F. & Scholl, H. Capture of Trojans by a growing proto-Jupiter. *Icarus* **131**, 41–51 (1998).
20. Henon, M. Numerical exploration of the restricted problem. VI. Hill's case: non-periodic orbits. *Astron. Astrophys.* **9**, 24–36 (1970).
21. Winter, O. C. & Vieira Neto, E. Time analysis for temporary gravitational capture: stable orbits. *Astron. Astrophys.* **377**, 1119–1127 (2001).
22. Saha, P. & Tremaine, S. The orbits of the retrograde Jovian satellites. *Icarus* **106**, 549–562 (1993).
23. Contopoulos, G. The "third" integral in the restricted three-body problem. *Astrophys. J.* **142**, 802–804 (1965).
24. Kozai, Y. Secular perturbations of asteroids with high inclinations and eccentricities. *Astron. J.* **67**, 591–598 (1962).
25. Goldreich, P., Lithwick, Y. & Sari, R. Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* **420**, 643–646 (2002).
26. Stern, S. A. & McKinnon, W. B. Triton's surface age and impactor population revisited in light of Kuiper Belt fluxes: Evidence for small Kuiper Belt objects and recent geological activity. *Astron. J.* **119**, 945–952 (2000).
27. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. *Numerical Recipes in C*, 2nd edn 724–732 (Cambridge Univ. Press, Cambridge, 1999).
28. Aarseth, S. From NBODY1 to NBODY6: The growth of an industry. *Publ. Astron. Soc. Pacif.* **111**, 1333–1346 (1999).
29. Brunello, A. F., Uzer, T. & Farrelly, D. Hydrogen atom in circularly polarized microwaves: Chaotic ionization via core scattering. *Phys. Rev. A* **55**, 3730–3745 (1997).
30. Lee, E., Brunello, A. F. & Farrelly, D. Coherent states in a Rydberg atom: Classical dynamics. *Phys. Rev. A* **55**, 2203–2221 (1997).

Acknowledgements This work was supported by the US National Science Foundation, the Royal Society (UK) and the US Office of Naval Research.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.W. (s.wiggins@bristol.ac.uk) or D.F. (david@habanero.chem.usu.edu).

A theory of power-law distributions in financial market fluctuations

Xavier Gabaix*, Parameswaran Gopikrishnan†‡, Vasiliki Plerou† & H. Eugene Stanley†

* Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

† Center for Polymer Studies and Department of Physics, Boston University, Boston, Massachusetts 02215, USA

Insights into the dynamics of a complex system are often gained by focusing on large fluctuations. For the financial system, huge databases now exist that facilitate the analysis of large fluctuations and the characterization of their statistical behaviour^{1,2}. Power laws appear to describe histograms of relevant financial fluctuations, such as fluctuations in stock price, trading volume and the number of trades^{3–10}. Surprisingly, the exponents that characterize these power laws are similar for different types and sizes of markets, for different market trends and even for different countries—suggesting that a generic theoretical basis may underlie these phenomena. Here we propose a model, based on a plausible set of assumptions, which provides an explanation for these empirical power laws. Our model is based on the hypothesis that large movements in stock market activity arise from the trades of large participants. Starting from an empirical characterization of the size distribution of those large market participants (mutual funds), we show that the power laws observed in financial data arise when the trading behaviour is performed in an optimal way. Our model additionally explains certain striking empirical regularities that describe the relationship between large fluctuations in prices, trading volume and the number of trades.

Define p_t as the price of a given stock and the stock price 'return' r_t as the change of the logarithm of stock price in a given time interval Δt , $r_t \equiv \ln p_t - \ln p_{t-\Delta t}$. The probability that a return has an absolute value larger than x is found empirically to be (see Fig. 1)^{4,8}:

$$P(|r_t| > x) \sim x^{-\zeta_r} \quad (1)$$

with $\zeta_r \approx 3$. Empirical studies also show that the distribution of trading volume V_t obeys a similar power law⁹:

$$P(V_t > x) \sim x^{-\zeta_V} \quad (2)$$

with $\zeta_V \approx 1.5$, while the number of trades N_t obeys¹⁰:

$$P(N_t > x) \sim x^{-\zeta_N} \quad (3)$$

with $\zeta_N \approx 3.4$.

The 'inverse cubic law' of equation (1) is rather 'universal', holding over as many as 80 standard deviations for some stock markets, with Δt ranging from one minute to one month, across different sizes of stocks, different time periods, and also for different stock market indices^{4,8}. Moreover, the most extreme events—including the 1929 and 1987 market crashes—conform to equation (1),

‡ Present address: Goldman Sachs and Co., 10 Hanover Square, New York, New York 10005, USA.

demonstrating that crashes do not appear to be outliers of the distribution. We test the universality of equations (2) and (3) by analysing the 35 million transactions of the 30 largest stocks on the Paris Bourse over the 5-yr period 1994–1999. Our analysis shows that the power laws (2) and (3) obtained for US stocks also hold for a distinctly different market, consistent with the possibility that equations (2) and (3) are as universal as equation (1).

Here, we develop a model that demonstrates how trading by large market participants explains the above power laws. We begin by noting that large market participants have large price impacts^{11–14}. To see why this is the case, observe that a typical stock has a turnover (fraction of shares exchanged) of approximately 50% a year, which implies a daily turnover of approximately 50%/250 = 0.2%—that is, on average 0.2% of outstanding shares change hands each day. The 30th-largest mutual fund owns about 0.1% of such a stock (Center for Research in Security Prices; <http://gsbwww.uchicago.edu/research/crsp/>). If the manager of such a fund sells its holdings of this stock, the sale will represent half of the daily turnover, and so will affect both the price and the total volume^{15–17}. Such a theory where large individual participants move the market is consistent with the evidence that stock market movements are difficult to explain with changes in fundamental values¹⁸.

Accordingly, we first perform an empirical analysis of the distribution of the largest market participants—mutual funds. We find, for each year of the period 1961–1999, that for the top 10% of distribution of the mutual funds, the market value of the managed assets S obeys the power law

$$P(S > x) \sim x^{-\zeta_S} \quad (4)$$

with $\zeta_S = 1.05 \pm 0.08$. Exponents of approximately one have also been found for the cumulative distributions city size¹⁹ and firm sizes^{20,21}, and the origins of this ‘Zipf’ distribution are becoming better understood²². On the basis of the assumption that managers of large funds trade on their beliefs about the future direction of the market, and that they adjust their speed of trading to avoid moving the market too much, we will see that their trading activity leads to $\zeta_r = 3$ and $\zeta_V = 1.5$.

In order to proceed, we first present empirical evidence for the shape of the price impact, then propose an explanation for this shape, and finally show how the resulting trading behaviour generates power laws (1)–(3).

First, the price impact Δp of a trade of size V has been established to have an increasing and concave functional form that is similar for

a large number of stocks^{23,24}. We hypothesize that for large volumes V its functional form is:

$$r = \Delta p \approx kV^{1/2} \quad (5)$$

for some constant k . So we investigate empirically the relation (Fig. 2):

$$E[r^2 | V] \sim V \quad (6)$$

which is supported by standard statistical tests. Because relation (5) implies $P(r > x) \sim P(kV^{1/2} > x) = P(V > x^2/k^2) \sim x^{-2\zeta_V}$, it follows that:

$$\zeta_r = 2\zeta_V \quad (7)$$

Thus, the power law of returns, equation (1), follows from the power law of volumes, equation (2), and the square-root form-price impact, equation (5). We next develop a framework for explaining equations (2) and (5).

We consider the behaviour of one stock whose original price is, say, one. The mutual fund manager who wishes to buy V shares offers a price increment Δp , so that the new price will become $1 + \Delta p$. Each seller i of size s_i who is offered a price increase Δp supplies the fund manager with q_i shares. Elementary considerations lead us to hypothesize $q_i \approx s_i \Delta p$ (see Supplementary Information). The number of sellers available after the fund manager has waited a time T is proportional to T . Thus after a time T , the fund manager can, on average, buy a quantity of shares equal to $kT\langle s \rangle \Delta p$ for some proportionality constant k . The search process stops (and the trades are executed simultaneously) when the desired quantity V is reached—that is, when $kT\langle s \rangle \Delta p = V$, so the time needed to find the shares is:

$$T = \frac{V}{\langle s \rangle k \Delta p} \sim \frac{V}{\Delta p} \quad (8)$$

Hence there is a trade-off between cost Δp and the time to execution T ; if the fund manager desires to realize the trade in a short amount of time T , the manager must pay a large price impact $\Delta p \sim V/T$.

Let us consider the fund manager’s decision problem. Managers trade on the assumption that a given stock is mispriced by an amount M , defined as the difference between the fair value of the stock and the traded price^{14,25,26}. The manager wants to exploit this mispricing quickly, as he expects that the mispricing will be progressively corrected, that is, expects that the price will increase

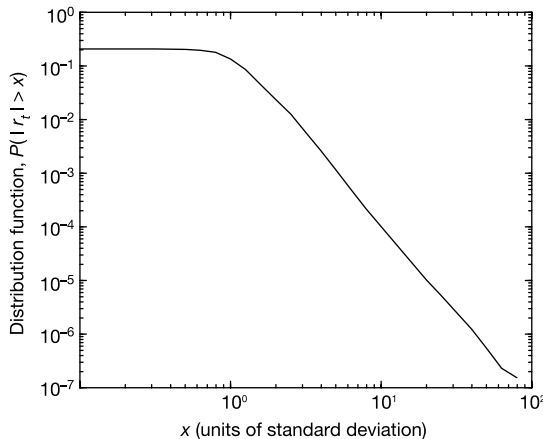


Figure 1 Cumulative distributions of the normalized 15-min absolute returns of the 1,000 largest companies in the ‘Trades and Quotes’ database for the 2-yr period 1994–1995. We define the normalized return as $r_{it} = (\tilde{r}_{it} - \tilde{r}_i)/\sigma_i$, where $\tilde{r}_i$ and σ_i are the mean and the standard deviation of the unnormalized return $\tilde{r}_{it}$ of stock i . We obtain $P(|r_t| > x) \sim x^{-\zeta_r}$ with $\zeta_r = 3.1 \pm 0.1$.

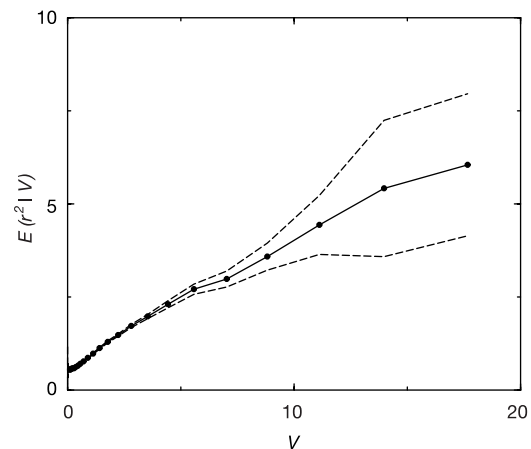


Figure 2 Conditional expectation of the squared return r^2 given the volume V . Here, the return is normalized as in Fig. 1, and the volume is normalized as $V_i = \tilde{V}_i/\tilde{V}_i$, where $\tilde{V}_i$ is the average of the unnormalized volumes $\tilde{V}_{it}$ of stock i . The bands represent 95% confidence intervals. The theory predicts a relation $E[r^2 | V] = aV + b$, the ‘square root’ price impact of volume. Statistical tests reported in the Supplementary Information confirm this relation.

at a rate μ . Hence, after a delay of T , the remaining mispricing is only $M - \mu T$. The total profit per share B/V is the realized excess return $M - \mu T$ minus the price concession Δp , which gives:

$$B = V(M - \mu T - \Delta p) \quad (9)$$

The fund manager's goal is thus to maximize B , the perceived dollar benefit from trading. The optimal price impact Δp maximizes B subject to equation (8), $T = aV/\Delta p$, that is, Δp maximizes $V(M - \mu aV/\Delta p - \Delta p)$, which gives equation (5).

The time to execution is $T \sim V/\Delta p \sim V^{1/2}$, and the number of 'chunks' in which the block is divided is $N \sim T \sim V^{1/2}$. These effects have been qualitatively documented in refs 11, 12, 23. The last relation gives:

$$\zeta_N = 2\zeta_V \quad (10)$$

which in turn predicts $\zeta_N = 3$, a value that is approximately consistent with the empirical value of 3.4 (ref. 10).

Thus far, we have a theoretical framework for understanding the square-root price impact of trades, equation (5), which with equation (2) explains the cubic law of returns, equation (1). We now focus on understanding equation (2).

We show that returns and volumes are power-law distributed with tail exponents:

$$\zeta_r = 3 \text{ and } \zeta_V = 3/2 \quad (11)$$

provided the following four conditions hold: (1) the power law exponent of mutual fund sizes is $\zeta_S = 1$ (Zipf's law); (2) the price impact follows the square root law, equation (5); (3) funds trade in typical volumes $V \approx S^\delta$ with $\delta > 0$; and (4) funds adjust trading frequency and/or volume so as to pay transactions costs in such a way that if we define $c(S)$ as:

$$c(S) \equiv \frac{\text{Annual amount lost by the fund in price impact}}{\text{Value } S \text{ of the assets under management}} \quad (12)$$

then $c(S)$ is independent of S for large S .

The empirical validity of conditions (1) and (2) was shown above, while condition (3) is a weak, largely technical, assumption discussed in the Supplementary Information. Condition (4) means that funds in the upper tail of the distribution pay roughly similar annual price-impact costs; that is, $c(S)$ reaches an asymptote for large sizes. We interpret this as an evolutionary 'survival constraint'. Funds that would have a very large $c(S)$ would have small returns and would be eliminated from the market. The average return $r(S)$ of funds of size S is independent of S (ref. 27). Because both small and large funds have similarly low ability to outperform the market, $c(S)$ is also independent of S .

For each block trade $V(S)$ a fund of size S incurs a price impact proportional to $V\Delta p$ which, from condition (2), is $V^{3/2}$. If $F(S)$ is the fund's annual frequency of trading, then the annual loss in transactions costs is $F(S) \cdot V^{3/2}$, so:

$$c(S) = F(S) \cdot [V(S)]^{3/2} / S \quad (13)$$

Condition (4) implies that either $V(S)$ or $F(S)$ will adjust in order to satisfy:

$$F(S) \sim S \cdot [V(S)]^{-3/2} \quad (14)$$

Condition (1) implies that the probability density function for mutual funds of size S is $\rho(S) \sim S^{-2}$. Because condition (3) states that $V \sim S^\delta > x$, and because they trade with frequency given by $F(S)$ in equation (14):

$$\begin{aligned} P(V > x) &\sim \int_{S^\delta > x} F(S) \rho(S) dS \\ &\sim \int_{S > x^{1/\delta}} S^{1-3\delta/2} S^{-2} dS \sim x^{-3/2} \end{aligned} \quad (15)$$

which leads to a power-law distribution of volumes with exponent

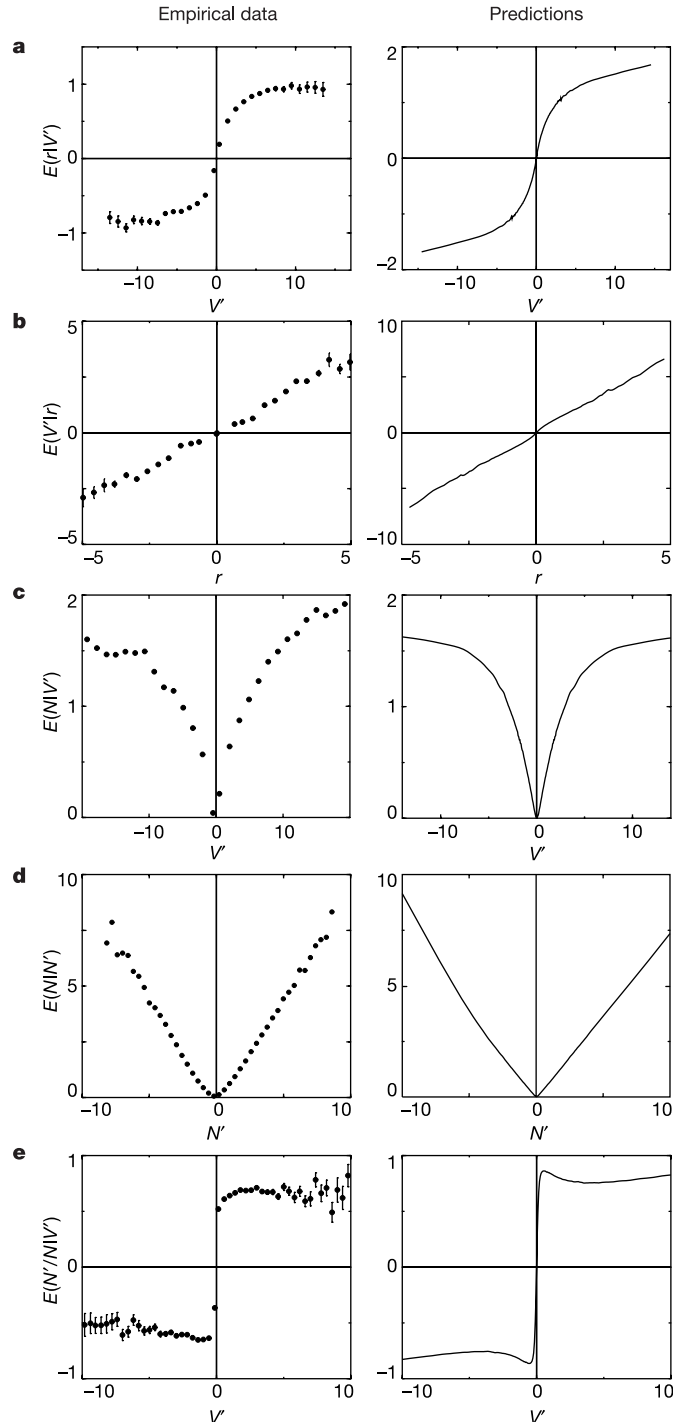


Figure 3 Conditional expectations for $E(r|V')$, $E(V'|r)$, $E(N|V')$, $E(N'|N')$, and $E(N'/N|V')$. We form, for each interval $\Delta t = 15$ min, the quantities (1) r , the return; (2) V_B (or V_S), the number of shares exchanged in a buyer- (or seller-) initiated trade²⁸; (3) N_B (or N_S), the number of buyer- (or seller-) initiated trades, $V' \equiv V_B - V_S$, and $N' \equiv N_B - N_S$. The left panels show the empirical values for the 116 most frequently traded stocks in the 'Trades and Quotes' database for the 2-yr period 1994–1995. Variables are normalized to unit variance after setting the mean to zero; for variables such as volume for which the variance does not exist, we have normalized by the first moment instead. The right panels show the model's predictions, which agree well with the empirical data.

$\zeta_V = 3/2$. Moreover, from equation (7), it follows that $\zeta_r = 3$. In addition, the above result does not depend on details of the trading strategy, such as the specific value of δ . (The Supplementary Information indicates a number of ways in which one can weaken the assumptions of independent and identical distributions made in this Letter.)

Although our model is mainly motivated by the regularities of returns, volume and number of trades taken separately, we also make predictions for the joint behaviour of those quantities. In a given time interval Δt , there will be J 'rounds' where a fund manager creates one or more trades. Each round j creates a volume V_j , a return $\pm V_j^{1/2}$ and a number of trades $V_j^{1/2}$. Then the total volume, number of trades, and returns, will be $V \equiv \sum_{j=1}^J V_j$, $N \equiv \sum_{j=1}^J V_j^{1/2}$ and $r \equiv \sum_{j=1}^J \epsilon_j V_j^{1/2}$, with $\epsilon_j = \pm 1$. As a measure of trade imbalance, we use N' , the number of buyer-initiated trades minus the number of seller-initiated trades²⁸, and V' , the number of shares exchanged in a buyer-initiated trade minus the number of shares exchanged in seller-initiated trades.

We next focus on equal-time relationships between V , N , and V' (ref. 24), using data from the 'Trades and Quotes' data base (New York Stock Exchange; <http://www.nyse.com>). These equal-time relationships are found to be universal across the large set of stocks analysed in ref. 24. Figure 3a shows that the prices impact function $E(r|V')$ produced by the model matches data. We observe that $J \gg 1$ (aggregation over several trades) flattens the shape of the price impact versus V . We study a variant of Fig. 3a in Fig. 3b, which plots $E(V'|r)$. Surprisingly, the shape is now roughly linear, a feature predicted by the model. The cause of the linearity is, again, the aggregation over several trades. Figure 3c, $E(N|V')$, tests the model prediction that periods with large volume imbalance V' are periods where a large number N of trades are made. One sees that the data display a relationship that is similar to that predicted by the model. Figures 3a–c support the view that large returns and large numbers of trades go together with large volume imbalances V' .

It is an important feature of the model that large trades beget more trades. Indeed, in our model:

$$|N'| \sim N \quad (16)$$

for large N and is dominated by one large fund manager who desires to trade a volume V_j , and creates a number of orders $V_j^{1/2}$, so that N_j , N , $|N'_j|$ and $|N'|$ have the same order of magnitude, $V_j^{1/2}$. Relation (16) means that most trades have the same sign, that is, move the price in the same direction—with the sign of the trade of the large fund manager. Equation (16) is indeed consistent with the empirical data shown in Fig. 3d. This contrasts with a simple alternative model where each desire to trade would create only one trade, as in a competitive market. In this alternative model we would have $N' = \sum_{i=1}^N \epsilon_i$, where $\epsilon_i = \pm 1$, leading to $|N'| \sim N^{1/2}$ in the tail events or $E[N|N'] \sim N^{1/2}$ in contrast to the data in Fig. 3d. Figure 3e supports the view that in periods of high volume imbalance V' , most trades change the price in the same direction. Indeed, the data and the model exhibit a similar sharp transition of N'/N as V' changes sign. $\square$

Received 10 December 2002; accepted 4 April 2003; doi:10.1038/nature01624.

1. Takayasu, H. (ed.) *Empirical Science of Financial Fluctuations: The Advent of Econophysics* (Springer, Berlin, 2002).
2. Bunde, A., Schellnhuber, H. J. & Kropp, J. (eds) *The Science of Disasters: Climate Disruptions, Heart Attacks, and Market Crashes* (Springer, Berlin, 2002).
3. Mandelbrot, B. B. The variation of certain speculative prices. *J. Business* **36**, 394–419 (1963).
4. Lux, T. The stable Paretian hypothesis and the frequency of large returns: an examination of major German stocks. *Appl. Fin. Econ.* **6**, 463–475 (1996).
5. Liu, Y. et al. The statistical properties of the volatility of price fluctuations. *Phys. Rev. E* **60**, 1390–1400 (1999).
6. Guillaume, D. M. et al. From the bird's eye to the microscope: a survey of new stylized facts of the intra-daily foreign exchange markets. *Fin. Stochastics* **1**, 95–129 (1997).
7. Plerou, V., Gopikrishnan, P., Amaral, L. A. N., Meyer, M. & Stanley, H. E. Scaling of the distribution of price fluctuations of individual companies. *Phys. Rev. E* **60**, 6519–6529 (1999).
8. Gopikrishnan, P., Plerou, V., Amaral, L. A. N., Meyer, M. & Stanley, H. E. Scaling of the distributions of fluctuations of financial market indices. *Phys. Rev. E* **60**, 5305–5316 (1999).

9. Gopikrishnan, P., Plerou, V., Gabaix, X. & Stanley, H. E. Statistical properties of share volume traded in financial markets. *Phys. Rev. E* **62**, R4493–R4496 (2000).
10. Plerou, V., Gopikrishnan, P., Amaral, L. A. N., Gabaix, X. & Stanley, H. E. Economic fluctuations and anomalous diffusion. *Phys. Rev. E* **62**, R3023–R3026 (2000).
11. Keim, D. & Madhavan, A. Transactions costs and investment style: An inter-exchange analysis of institutional equity trades. *J. Fin. Econ.* **46**, 265–292 (1997).
12. Chan, L. & Lakonishok, J. The behavior of stock prices around institutional trades. *J. Fin.* **50**, 1147–1174 (1995).
13. Wurgler, J. & Zhuravskaya, K. Does arbitrage flatten demand curves for stocks? *J. Business* **75**, 583–608 (2002).
14. Bagwell, L. Dutch auction repurchases: An analysis of shareholder heterogeneity. *J. Fin.* **47**, 71–105 (1992).
15. Kyle, A. S. Continuous auctions and insider trading. *Econometrica* **53**, 1315–1335 (1985).
16. Grossman, S. & Miller, M. Liquidity and market structure. *J. Fin.* **43**, 617–633 (1988).
17. O'Hara, M. *Market Microstructure Theory* (Blackwell, Oxford, 1997).
18. Cutler, D., Poterba, J. M. & Summers, L. H. What moves stock prices? *J. Portfolio Management* **15**, 4–12 (1989).
19. Kyle, G. *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, 1949).
20. Okuyama, K., Takayasu, M. & Takayasu, H. Zipf's law in income distribution of companies. *Physica A* **269**, 125–131 (1999).
21. Axtell, R. Zipf distribution of U.S. firm sizes. *Science* **293**, 1818–1820 (2001).
22. Gabaix, X. Zipf's law for cities: An explanation. *Q. J. Econ.* **114**, 739–767 (1999).
23. Hasbrouck, J. Measuring the information content of stock trades. *J. Fin.* **46**, 179–207 (1991).
24. Plerou, V., Gopikrishnan, P., Gabaix, X. & Stanley, H. E. Quantifying stock price response to demand fluctuations. *Phys. Rev. E* **66**, 027104 (2002).
25. Daniel, K., Hirshleifer, D. & Subrahmanyam, A. Investor psychology and security market under- and over-reactions. *J. Fin.* **53**, 1839–1885 (1998).
26. Shleifer, A. *Inefficient Markets: An Introduction to Behavioral Finance* (Oxford Univ Press, Oxford, 2000).
27. Gabaix, X., Ramalho, R. & Reuter, J. Power laws and mutual fund dynamics (MIT Mimeo, Massachusetts Institute of Technology, Cambridge, 2003).
28. Lee, C. M. C. & Ready, M. J. Inferring trade direction from intraday data. *J. Fin.* **46**, 733–746 (1991).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the National Science Foundation's economics program and the Russell Sage Foundation for support, K. Doran for research assistance, and M. Avellaneda, R. Barro, O. Blanchard, J. Campbell, A. Dixit, J. Hasbrouck, C. Hopman, D. Laibson, L. Pedersen, T. Philippon, R. Ramalho, J. Reuter, G. Saar, J. Scheinkman, A. Shleifer, D. Vayanos and J. Wang for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.G. (xgabaix@mit.edu).

Measurement of the displacement field of dislocations to 0.03 Å by electron microscopy

Martin J. Hÿtch*, Jean-Luc Putaux† & Jean-Michel Pénisson‡

* Centre d'Etudes de Chimie Métallurgique, Centre National de Recherche Scientifique (CNRS), 15 rue G. Urbain, 94407 Vitry-sur-Seine, France

† Centre de Recherches sur les Macromolécules Végétales, Centre National de Recherche Scientifique (CNRS), BP 53, 38041 Grenoble, France; Joseph Fourier University, BP 53, 38041 Grenoble, France

‡ Département de Recherche Fondamentale sur la Matière Condensée (DRFMC/SP2M/ME) CEA-Grenoble, 17 Avenue des Martyrs, 38054 Grenoble, France

Defects and their associated long-range strain fields are of considerable importance in many areas of materials science^{1,2}. For example, a major challenge facing the semiconductor industry is to understand the influence of defects on device operation, a task made difficult by the fact that their interactions with charge carriers can occur far from defect cores, where the influence of the defect is subtle and difficult to quantify^{3,4}. The accurate measurement of strain around defects would therefore allow more detailed understanding of how strain fields affect small structures—in particular their electronic, mechanical and chemi-

cal properties—and how such fields are modified when confined to nanometre-sized volumes. Here we report the measurement of displacements around an edge dislocation in silicon using a combination of high-resolution electron microscopy and image analysis inherited from optical interferometry. The agreement of our observations with anisotropic elastic theory calculations is better than 0.03 Å. Indeed, the results can be considered as an experimental verification of anisotropic theory at the near-atomic scale. With the development of nanostructured materials and devices, we expect the use of electron microscopy as a metrological tool for strain analysis to become of increasing importance.

High-resolution electron microscopy is a likely candidate for atomic-scale strain mapping, as images are formed of the atomic lattice. In principle, therefore, strain and local deformation can be determined directly by measuring the displacement of the lattice fringes in the image. However, there are a number of difficulties that have to be overcome before this can be achieved with sufficient accuracy and reliability. First, the displacement of lattice fringes in a high-resolution image do not correspond exactly to those of the atomic lattice. Recent theoretical predictions have nevertheless shown that for small strains and centrosymmetric crystals (as here) the correspondence is almost exact⁵. Second, specimen preparation has to be optimized to provide images with sufficiently high signal-to-noise ratios and to avoid undue strain relaxation and buckling⁶. Last, computational methods have had to be developed in order to extract the information from images, and to eliminate the optical distortions introduced by the microscope itself.

Different approaches have already been applied to the measurement of localized displacements using electron microscopy. (1) Displacements of individual atomic columns at interfaces have been determined to 0.1 Å accuracy by comparing image simulations with experimental images⁷. (2) Twin boundary expansions have been measured to 0.04 Å by elimination of lens aberrations by focal reconstruction⁸. (3) A combination of high-resolution electron microscopy and selected-area electron diffraction has been used to determine the positions of atoms in the unit cell for perfect crystals to 0.02 Å (ref. 9). Error analysis in these cases, even the most rigorous⁷, concerned only the measurement itself.

In order to establish high-resolution electron microscopy as a metrological tool for strain analysis at the atomic level, it is therefore essential to determine the displacement field for a known object and to compare the results directly with theory. Here we propose a method for measuring continuous displacement fields around defects using the geometric phase technique¹⁰, and apply the

analysis to a pure edge dislocation—a textbook case for elastic theory owing to the fundamental role played by dislocations in plasticity¹¹. We will show that experimental measurements and theory agree to within 0.03 Å.

Mapping displacement fields using high-resolution electron microscopy was first proposed¹² for the study of strained semiconductor multilayers. The method we use is an extension of a technique first developed in optical interferometry¹³, and later proposed for the study of defects¹⁴. We have aimed to obtain the highest accuracy possible by optimizing specimen preparation, and by introducing correction of the optical distortions due to the projector lenses of the electron microscope.

The geometric phase approach is based on combining real-space and Fourier-space information, an idea first proposed for the analysis of sound in terms of wave packets¹⁵. Displacements are measured by calculating the 'local' Fourier components of the lattice fringes in an image¹⁶. In this formulation, the intensity in an image, $I(\mathbf{r})$, is written as:

$$I(\mathbf{r}) = \sum_{\mathbf{g}} I_{\mathbf{g}}(\mathbf{r}) e^{2\pi i \mathbf{g} \cdot \mathbf{r}} \quad (1)$$

where $\mathbf{g}$ are the reciprocal lattice vectors describing the undistorted lattice. The local Fourier components are obtained by filtering in Fourier space, and have an amplitude and phase:

$$I_{\mathbf{g}}(\mathbf{r}) = A_{\mathbf{g}}(\mathbf{r}) e^{i P_{\mathbf{g}}(\mathbf{r})} \quad (2)$$

where the amplitude, $A_{\mathbf{g}}(\mathbf{r})$, describes the local contrast of the fringes and the phase, $P_{\mathbf{g}}(\mathbf{r})$, their position. The phase is related simply to the displacement field $\mathbf{u}(\mathbf{r})$ by the following relation:

$$P_{\mathbf{g}}(\mathbf{r}) = -2\pi \mathbf{g} \cdot \mathbf{u}(\mathbf{r}) \quad (3)$$

and by measuring two phase images, $P_{\mathbf{g}_1}(\mathbf{r})$ and $P_{\mathbf{g}_2}(\mathbf{r})$, the two-dimensional displacement field, can be determined¹⁰:

$$\mathbf{u}(\mathbf{r}) = -\frac{1}{2\pi} [P_{\mathbf{g}_1}(\mathbf{r}) \mathbf{a}_1 + P_{\mathbf{g}_2}(\mathbf{r}) \mathbf{a}_2] \quad (4)$$

Here $\mathbf{a}_1$ and $\mathbf{a}_2$ are the basis vectors for the lattice in real space corresponding to the reciprocal lattice defined by $\mathbf{g}_1$ and $\mathbf{g}_2$.

For the particular case of dislocations, analysis in terms of the phase can be developed from the general definition of the Burgers vector $\mathbf{b}$:

$$\mathbf{b} = \oint_L \nabla \mathbf{u} \cdot d\mathbf{l} \quad (5)$$

where L is the closed loop around the defect. Substitution of

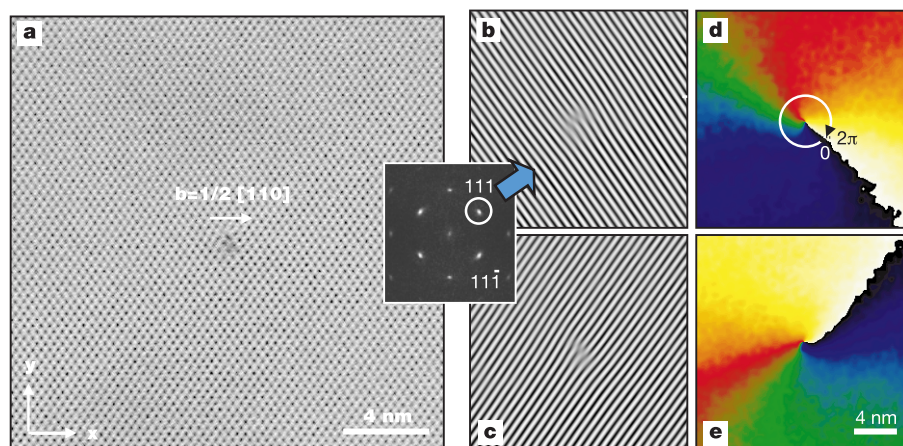


Figure 1 Geometric phase analysis of an edge dislocation seen end-on in silicon. **a**, High-resolution electron microscope image in $[1\bar{1}0]$ orientation, Burgers vector $\mathbf{b} = 1/2[110]$, Fourier transform inset; **b**, (111) lattice fringes obtained by filtering (magnified for display

purposes); **c**, $(11\bar{1})$ lattice fringes; **d**, phase image of (111) lattice fringes; **e**, phase image of $(11\bar{1})$ lattice fringes. Colour range 0 to 2π rad.

equation (4) for the displacement field results in the following relation:

$$\mathbf{b} = -\frac{1}{2\pi} \left[\mathbf{a}_1 \oint_L \nabla P_{g1} \cdot d\mathbf{l} + \mathbf{a}_2 \oint_L \nabla P_{g2} \cdot d\mathbf{l} \right] \quad (6)$$

which shows that discontinuities will be observed in the phase. A numerical way of determining their value from phase images has been proposed¹⁷. If the phase discontinuities are of the form $2n\pi$ on going clockwise around the dislocation core, the Burgers vector will be:

$$\mathbf{b} = n_1 \mathbf{a}_1 + n_2 \mathbf{a}_2 \quad (7)$$

the simplicity of which attests the naturalness of the phase description.

Electron microscopy was carried out on specially prepared silicon bicrystals¹⁸. Thin foils were prepared by mechanical polishing to a thickness of about 70 μm followed by ion milling in a Gatan DuoMill at 6 kV. The amorphous surface layer was removed by chemical etching in a HF 10%-HNO₃ 90% solution at 0 °C. Image processing was carried out using routines written for the software package Digital Micrograph (Gatan). Distortions of the displacement field due to the projector lenses were corrected to better than 0.1% across the whole field of view, using images of perfect crystal taken under identical experimental conditions.

Figure 1a shows a high-resolution image in $[1\bar{1}0]$ orientation of a dislocation seen end-on in silicon, taken on a JEOL 200CX operating at 200 kV (spherical aberration = 1.1 mm, point resolution 0.22 nm). Phase images were calculated for the two sets of $\{111\}$ lattice fringes (Fig. 1b, c) using gaussian masks, and are shown in Fig. 1d, e. The basis vectors for the phase calculation are the following: $\mathbf{g}_1 = [11\bar{1}]^*$, $\mathbf{g}_2 = [111]^*$ and $\mathbf{a}_1 = 1/4[11\bar{2}]$, $\mathbf{a}_2 = 1/4[112]$, where an asterisk indicates reciprocal space coordinates. The phase increases monotonically around the dislocation cores, with an abrupt change from 0 to 2π radians owing to the normalization of the phase in the range of 2π . In both cases the phase discontinuities, n_1 and n_2 , are unity, indicating one extra $(11\bar{1})$ and (111) plane and therefore: $\mathbf{b} = \mathbf{a}_1 + \mathbf{a}_2 = 1/4[11\bar{2}] + 1/4[112] = 1/2[110]$ corresponding to a pure edge dislocation—the Lomer dislocation.

Taking the x axis parallel to $[220]$ and the y axis parallel to $[002]$,

the displacement field $\mathbf{u} = (u_x, u_y)$, calculated from equation (4), is shown in Fig. 2. (The double discontinuity is due to the fact the largest lattice spacings in the x direction are the (220) lattice planes for which $n = 2$.) Also shown in Fig. 2 is the displacement field calculated using anisotropic inelastic theory with the standard elastic constants for silicon¹¹. Qualitatively it can be seen that the agreement is excellent. In order to analyse quantitatively the results and because of the complexity of the equations, it is useful to consider the result as given by isotropic elastic theory:

$$u_x = \frac{b}{2\pi} \left(\theta + \frac{\sin 2\theta}{4(1-\nu)} \right) \quad (8)$$

$$u_y = -\frac{b}{2\pi} \left(\frac{1-2\nu}{4(1-\nu)} \ln r^2 + \frac{\cos 2\theta}{4(1-\nu)} \right) \quad (9)$$

where r and θ are the polar coordinates centred on the core position, and ν the Poisson's constant. The first term in u_x describes the discontinuity in the displacement field, and increases linearly in θ ; it is the main contribution seen in the calculated and experimental displacement fields. The displacement in u_y has again two terms, this time one depending uniquely on r and the other in θ . We shall now consider in detail these sinusoidal variations in u_x and u_y .

Figure 3 shows the sinusoidal contributions to the displacement field, obtained by subtracting the theoretical first terms in each case. The variation in approximately $\sin 2\theta$ and $\cos 2\theta$ can be seen clearly. For a quantitative comparison with theory, the variation has been measured around a circle of radius 7.5 nm averaged over 1 nm radially for u_x and u_y (Fig. 4). It is here that we see clearly the effect of crystal anisotropy: the amplitude of the variations in the $[110]$ and $[001]$ directions are not the same. Isotropic theory is therefore not sufficient to describe the experimental displacement field measured here. Another detail is well reproduced—the sinusoidal variations lean slightly towards the smaller angles, that is, the positive slope is steeper than the negative. This is due to a small term in $\sin 4\theta$ in the displacement field due to the anisotropy but difficult to extract analytically from the equations. The agreement of the measured displacement field with the anisotropic theory (solid curves in Fig. 4) is striking, with a root-mean-square deviation of 0.03 Å, and no fitting parameters.

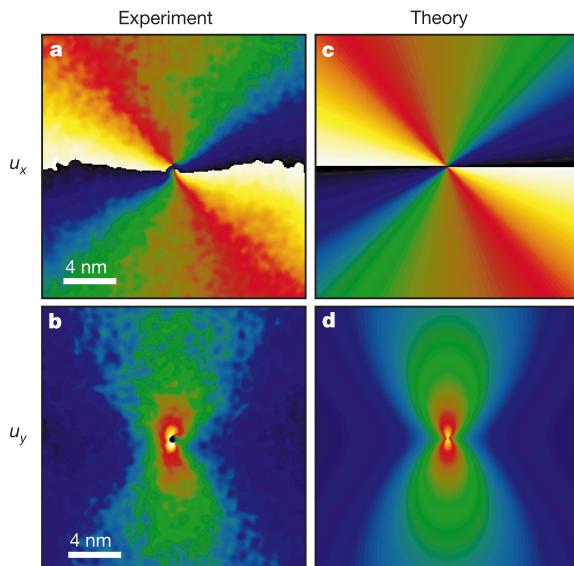


Figure 2 Experimental and theoretical displacement fields. Displacement field $\mathbf{u} = (u_x, u_y)$ calculated from phase images (a, b) and anisotropic elastic theory (c, d). Colour range for u_x (a, c) is 0 nm to 0.192 nm, and for u_y (b, d) is -0.271 nm to 0 nm (lattice spacings $d_{002} = 0.271$ nm, $d_{220} = 0.192$ nm).

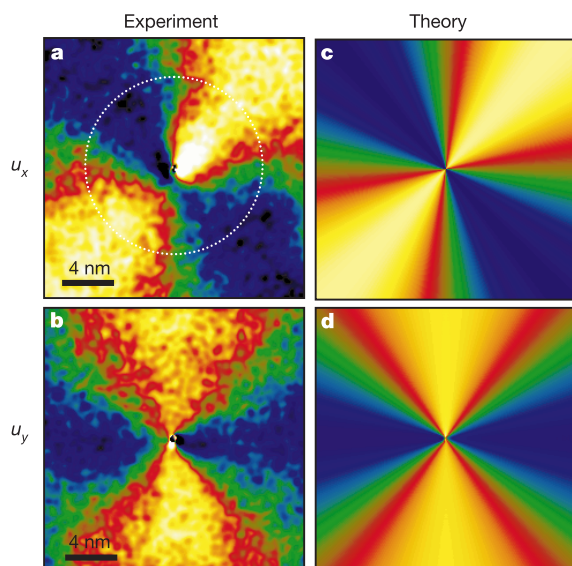


Figure 3 Sinusoidal component of the displacement field. Experimental and theoretical displacement field for u_x (a, c) and u_y (b, d), respectively. Colour range ± 0.3 Å. Circle in a shows location of measurements reported in Fig. 4.

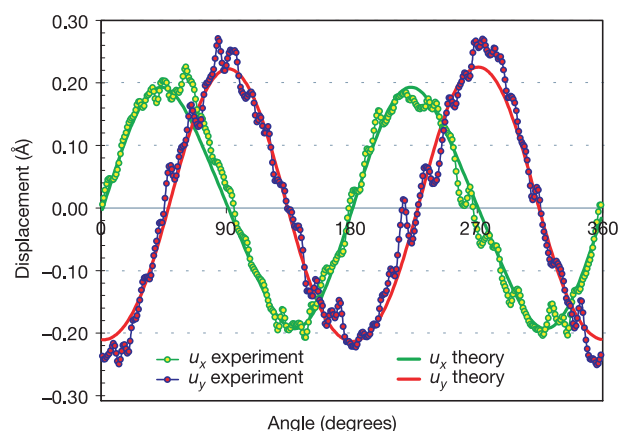


Figure 4 Angular variation of displacement field. Sinusoidal variation at radius 7.5 nm from core (see circle on Fig. 3a), measured experimentally and from anisotropic theory. Root-mean-square deviation from theoretical values is 0.03 Å (3 pm).

To further quantify these measurements, we have determined the amplitude of the sinusoidal variations for radii between 5 nm and 10 nm from the core and for two values of defocus (Table 1). The final experimental result is taken as the average for the two defoci, and the error their standard deviation. It can be seen that the agreement is within error bars of 0.005 Å, or 0.5 pm. As an extra check on results, the two defoci taken separately agree to within the experimental error. The amplitude difference between u_x and u_y predicted by anisotropic theory is only 3.4 pm. An accuracy of better than 1 pm is therefore necessary to distinguish with confidence isotropic from anisotropic elastic theory.

If we assume now that the elastic constants of silicon are unknown, we can obtain an estimate from the experimental results. In silicon there are just three independent coefficients (c_{11} , c_{12} and c_{44}) but only two measured values (amplitudes in the [110] and [001] directions). If we fix the value of one coefficient arbitrarily, we obtain a precision of 5–10% for the other two, and a precision of about 20% for the coefficient of anisotropy H . We could imagine measuring the displacement field for a dislocation in a different orientation to reduce the number of degrees of freedom. For non-cubic crystals, measurements would at least give an estimation of the degree of crystal anisotropy.

We have shown that continuous displacement fields can be measured at the nanometre scale to an accuracy of 0.03 Å, assuming anisotropic elastic theory to be correct. Indeed, predictions concerning the amplitude of displacements agree to better than 0.01 Å, 100 times the resolution of the electron microscope used. With this accuracy it was possible to distinguish between isotropic and anisotropic elastic theory. The most important factors are sample preparation, calibration of optical distortions and the resulting signal-to-noise ratio of the image. Thin film relaxation⁶ was not a factor, as no significant modification of the displacement field was observed as a function of distance from the core (in the region analysed). The possibility now exists of extending the analysis to the core region of dislocations to test the limits of elastic theory and of microscope imaging. More generally, the approach should allow the

Table 1 Experimental and theoretical displacement amplitudes

	u_x amplitude	u_y amplitude
Experiment	18.3 pm $\pm$ 0.5 pm	21.9 pm $\pm$ 0.4 pm
Theory	18.6 pm	22.0 pm

Top row, the mean and standard deviation of the amplitude of displacements around the dislocation core (see Fig. 4) measured for radii between 5 and 10 nm. Agreement with anisotropic elastic theory (bottom row) is to within 0.005 Å.

experimental validation of atomistic modelling of strain, which is being used increasingly for devices and materials at the nanometre scale. □

Received 23 October 2002; accepted 19 March 2003; doi:10.1038/nature01638.

1. Nabarro, F. R. N. (ed.) *Dislocations in Solids* Vols 1–11 (Amsterdam, North-Holland, 1979–2003).
2. Lépinoux, J., Mazière, D., Pontikis, V. & Saada, G. (eds) *Multiscale Phenomena in Plasticity: From Experiments to Phenomenology, Modelling and Materials Engineering* (NATO Science Series Applied Science Vol. 367, Kluwer Academic, Dordrecht, The Netherlands, 2000).
3. Semiconductor Industry Association *The National Technology Road Map for Semiconductors* at (<http://public.itrs.net/>) (2002).
4. Alexander, H. & Teichler, H. in *Handbook of Semiconductor Technology* (eds Jackson, K. A. & Schröter, W.) 293–376 (Wiley-VCH, Berlin, 2000).
5. Hytch, M. J. & Plamann, T. Reliable conditions for the measurement of displacement and strain from high resolution electron microscope images. *Ultramicroscopy* **87**, 199–212 (2001).
6. Treacy, M. M. J., Gibson, J. M. & Howie, A. On elastic relaxation and long wavelength microstructures in spinodally decomposed $\text{In}_x\text{Ga}_{1-x}\text{As}_y\text{P}_{1-y}$ epitaxial layers. *Phil. Mag. A* **51**, 389–417 (1985).
7. Möbus, G. & Kienle, O. Probability calculus for quantitative HREM. *Ultramicroscopy* **85**, 183–198 (2000).
8. Jia, C. L. & Thust, A. Investigation of atomic displacements at a $\Sigma 3$ [111] twin boundary in BaTiO_3 by means of phase retrieval electron microscopy. *Phys. Rev. Lett.* **82**, 5052–5055 (1999).
9. Weirich, T. E., Ramlau, R., Simon, A., Hovmöller, S. & Zou, X. A crystal structure determined with 0.02 Å accuracy by electron microscopy. *Nature* **382**, 144–146 (1996).
10. Hytch, M. J., Snoeck, E. & Kilaas, R. Quantitative measurement of displacement and strain fields from HREM micrographs. *Ultramicroscopy* **74**, 131–146 (1998).
11. Hirth, J. P. & Lothe, J. *Theory of Dislocations* (McGraw-Hill, New York, 1968).
12. Bierwolf, R. et al. Direct measurement of local lattice distortions in strained layer structures by HREM. *Ultramicroscopy* **49**, 273–285 (1993).
13. Takeda, M., Ina, H. & Kobayashi, S. Fourier-transform method of fringe pattern analysis for computer-based topography and interferometry. *J. Opt. Soc. Am.* **72**, 156–160 (1982).
14. Takeda, M. & Suzuki, J. Crystallographic heterodyne phase detection for highly sensitive lattice-distortion measurements. *J. Opt. Soc. Am. A* **13**, 1495–1500 (1996).
15. Gabor, D. Theory of communication. *Proc. Inst. Electr. Eng.* **93**, 429–457 (1946).
16. Hytch, M. J. Analysis of variations in structure from high resolution electron microscope images by combining real space and Fourier space information. *Microsc. Microanal. Microstruct.* **8**, 41–57 (1997).
17. Kret, S., Dluzewski, P., Dluzewski, P. & Sobczak, E. Measurement of dislocation core distribution by digital processing of high-resolution transmission electron microscopy micrographs: a new technique for studying defects. *J. Phys. C* **12**, 10313–10318 (2000).
18. Thibault-Desseaux, J., Putaux, J. L., Bourret, A. & Kirchner, H. O. K. Dislocations stopped by the $\Sigma(122)$ grain boundary in Si. An HREM study of thermal activation. *J. Phys. France* **50**, 2525–2540 (1989).

Acknowledgements This work was carried out as part of a CNRS funded European network (GDR-E) “Quantification and measurement in transmission electron microscopy”, involving laboratories in France, the UK, Germany and Switzerland. M.J.H. would like to thank M. Nowak for help with the elastic theory calculations.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.J.H. (martin.hytch@glvt-cnrs.fr).

Detection of bromine monoxide in a volcanic plume

N. Bobrowski*, G. Hönninger*†, B. Galle‡ & U. Platt*

* Institut für Umweltphysik, University of Heidelberg, INF 229, D69120 Heidelberg, Germany

‡ Chalmers University of Technology (CTH), S-41296 Gothenburg, Sweden

The emission of volcanic gases usually precedes eruptive activity¹, providing both a warning signal and an indication of the nature of the lava soon to be erupted. Additionally, volcanic emissions are a significant source of gases and particles to the atmosphere, influencing tropospheric and stratospheric trace-gas budgets². Despite some halogen species having been measured in volcanic plumes³ (mainly HCl and HF), little is known about bromine compounds⁴ and, in particular, gas-phase

† Present address: Meteorological Service of Canada, Toronto, Ontario M3H 5T4, Canada.

reactive bromine species. Such species are especially important in the stratosphere⁵, as reactive bromine—despite being two orders of magnitude less abundant than chlorine—accounts for about one-third of halogen-catalysed ozone depletion⁶. In the troposphere, bromine-catalysed complete ozone destruction has been observed to occur regularly during spring in the polar boundary layers^{7–11} as well as in the troposphere above the Dead Sea basin¹². Here we report observations of BrO and SO₂ abundances in the plume of the Soufrière Hills volcano (Montserrat) in May 2002 by ground-based multi-axis differential optical absorption spectroscopy. Our estimate of BrO emission leads us to conclude that local ozone depletion and small ozone ‘holes’ may occur in the vicinity of active volcanoes, and that the amount of bromine emitted from volcanoes might be sufficiently large to play a role not only in the stratosphere, but also in tropospheric chemistry.

A serious problem in volcanic research has been the difficult accessibility of the volcanic centres and the dangers of working there. An optical remote-sensing method is therefore of great advantage. Beginning with early developments marked by the COSPEC instrument¹³, ultraviolet—and infrared—spectroscopy has been used for the measurements of volcanic gases^{3,14}. Differential optical absorption spectroscopy (DOAS; see, for example, ref. 15) allows highly sensitive and specific detection of many trace gases (specific molecules, not just elements) using artificial light sources or scattered sunlight. An additional advantage of the scattered sunlight multi-axis (MAX)-DOAS technique used here^{11,16} is the possibility of probing the plume at various distances downwind of a volcanic source; this helps the understanding of the atmospheric chemistry involved, as well as the transport processes

responsible for plume dispersal. All this can be achieved with very compact, lightweight, portable, and relatively simple hardware¹⁴.

The island of Montserrat is located in the northern part of the Lesser Antilles in the Caribbean Sea (16.72° N, 62.18° W). The small volcanic island (Fig. 1) extends 16 km from north to south and 10 km from east to west. There are three volcanic centres classified as andesitic stratovolcanoes (<http://www.mvo.ms/>). The current eruption of the Soufrière Hills volcano began on 18 July 1995 (ref. 17). It has been characterized by the growth of an andesite lava dome with associated pyroclastic flows, vulcanian explosion and debris flows, similar to the activity that produced the ancient deposits making up the rest of the island.

We used a portable sequential scanning MAX-DOAS instrument^{11,15} to identify and quantify the atmospheric trace gases by their specific, structured absorption features in the ultraviolet spectral region. The fundamental equation of absorption spectroscopy is the Beer–Lambert law, which quantifies the attenuated radiation intensity $I(\lambda, L)$ upon passing through matter: $I(\lambda, L) = I_0(\lambda) \exp[-S(L)\sigma]$. Here $I_0(\lambda)$ denotes the radiation intensity (at wavelength λ) before passing the light path L in the atmosphere, $\sigma(\lambda, T)$ is the wavelength- and temperature (T)-dependent trace-gas absorption cross-section, and S (given by $S = \int_0^L c(l) dl$) is the slant column density (SCD) of the trace gas with concentration $c(l)$ as a function of position l along the light path. SCDs are directly determined by DOAS measurements. However, fluxes can be calculated using further estimates (see below). The complete measurement procedure, including initial data evaluation, was controlled by the software package DOASIS¹⁸ running on a notebook computer.

We took measurements from 23 to 31 May 2002 at two locations: (1) on 25 May, the measurements were taken from inside the daytime entry zone ~4 km to the east of the summit (LL in

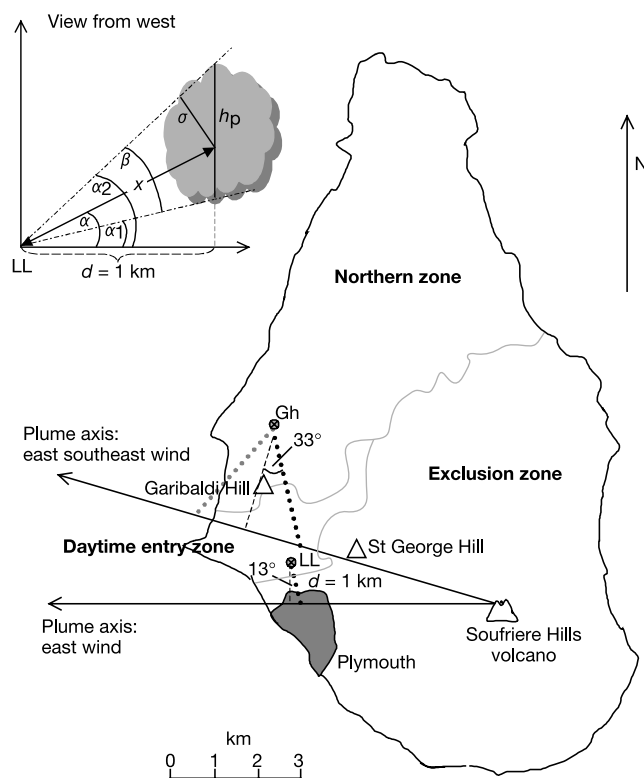


Figure 1 Study area. Montserrat island is divided into three zones; the exclusion zone, the daytime entry zone, and the northern zone. The two measurement sites are shown: LL (first site, near Lovers Lane, the measurement place on 25 May 2002), and Gh (second site, ‘Glenn’s house’, the measurement place for all days except 25 May). Dotted lines show the three different viewing directions. The suggested plume dispersion is also shown. Inset shows viewing geometry for 25 May 2002, suggested plume shape, and geometry as follows. $x = d/\cos(\alpha) \approx 1.2$ km, $\beta = \alpha_2 - \alpha_1$, $x \tan(\beta/2) \sigma$. Correction of observation azimuth: extension $\approx 1/\cos 13^\circ \approx 1.03$.

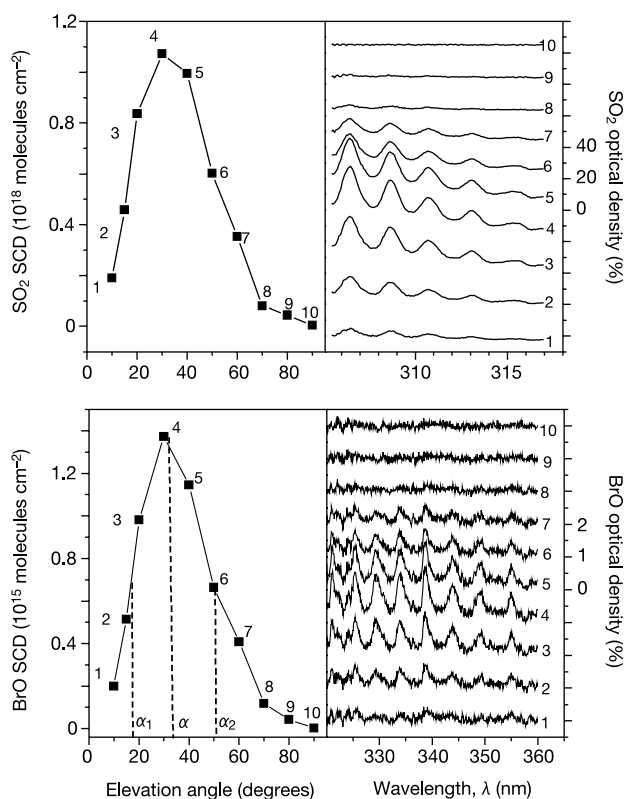


Figure 2 Cross-section of the plume from the observation point ~4 km from the summit, on 25 May 2002. The plume was observed over an angular range of 34°. Measured BrO and SO₂ spectra corresponding to the individual viewing directions are shown on the right. (In both panels, spectra—except spectrum 5—are vertically shifted for clarity).

Fig. 1); (2) on all other days, the instrument was located 6 km northwest from the volcanic summit just outside the exclusion zone (Gh in Fig. 1). The azimuthal viewing direction before 26 May was south-southwest to southeast, afterwards south-southwest (Fig. 1). On 24 May various viewing directions were tested, and thus the plume could be probed at different distances from the summit. Wind data were obtained from a weather station on the top of St George's Hill (350 m above sea level; Fig. 1). Telescope elevation angles α of 10°, 15°, 20°, 30°, 40°, 50°, 60°, 70°, 80° and 90° above the horizon were applied to scan the volcanic plume. Each scan sequence lasted about 10–20 minutes.

Both BrO and SO₂ could always be detected in each scan through the plume. SCD values ranged from near zero looking away from the plume, to 2×10^{15} molecules cm⁻² for BrO and $>1.7 \times 10^{18}$ molecules cm⁻² for SO₂, respectively, in the centre of the plume scan (Fig. 2). The SCD maxima in the plume varied by a factor of up to three during the day. In Fig. 3 the BrO and SO₂ SCDs for a series of plume scans during 24 May 2002 are shown. The general pattern shows the behaviour expected for the measurement routine. For each MAX-DOAS series (3–8 per hour), the maximum SCD corresponds to the viewing direction towards the centre of the plume, while lower SCDs are measured for viewing directions (elevation) above or below the plume (see Fig. 1 inset). SCDs approach zero for the highest and lowest elevation angles, showing that the plume was well captured. In the morning, the viewing azimuth direction was towards the summit of the Soufrière Hills volcano (black dotted line in Fig. 1)—that is, closer to the source compared to the afternoon situation, where the viewing direction pointed away from the summit (grey dotted line in Fig. 1). SO₂ was closely correlated with BrO (Fig. 4). The BrO/SO₂ molar ratio of 8.2×10^{-4} corresponds to a Br/S mass ratio of $\sim 2 \times 10^{-3}$.

From a simple geometric approach (triangulation, see Fig. 1), we can infer the approximate extension of the plume (perpendicular to its propagation direction) based solely on the MAX-DOAS measurements and wind direction. For example, on 25 May (Fig. 2), the BrO plume width (full-width at half-maximum, FWHM) seen from the instrument location ranged over 34°. The distance (d), measured perpendicular, from the plume centre to the observation point was estimated at ~ 1 km using the wind direction data from the local weather station. The corresponding plume width (see equations in Fig. 1 legend) is ~ 760 m (2σ). In comparison a simple dispersion model (assuming a wind speed of 10 m s^{-1} , an emission height of 1,000 m above sea level, a turbulent diffusion coefficient of $K \approx 100 \text{ m}^2 \text{ s}^{-1}$, and $\sigma = (2Kt)^{1/2}$) yields a very similar plume diameter of $2\sigma \approx 600$ m at 4 km ($t = 400$ s transport time) downwind of the source. Combined with the measured SCDs

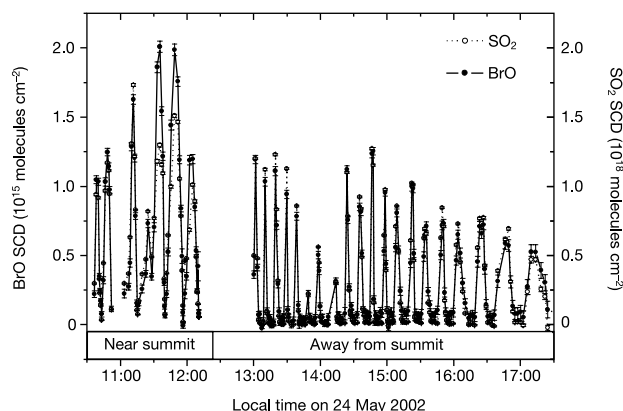


Figure 3 Series scans of the volcanic plume on 24 May 2002, showing slant column densities of BrO and SO₂. The viewing azimuth in the morning was towards the Soufrière Hills summit, and in the afternoon away from the summit.

(BrO $\approx 2 \times 10^{15}$ molecules cm⁻², SO₂ $\approx 1.75 \times 10^{18}$ molecules cm⁻²), we derive mixing ratios of ~ 1 part per billion (p.p.b.) for BrO and ~ 1 p.p.m. for SO₂, respectively, in the plume.

Using our SCD measurements, we can estimate the total SO₂ and BrO source strengths of the Soufrière Hills volcano. Assuming that the plume has rotational symmetry around its propagation direction, we calculate a BrO area density ($A \approx S2\sigma\pi/4$) of $\sim 8.4 \times 10^{19}$ molecules cm⁻¹ for the scan shown in Fig. 2. Including the measured wind speed as a lower limit for the total wind speed ($v = 10 \text{ m s}^{-1}$), we estimate a mass flow rate $\Phi(\text{BrO}) = Av$ of 8.4×10^{22} molecules s⁻¹ BrO or 11.1 g s^{-1} Br (~ 350 t reactive Br per year). Analogous calculations for SO₂ yield a total source strength of $\sim 5.6 \text{ kg S s}^{-1}$ ($1.8 \times 10^5 \text{ t S yr}^{-1}$).

For a global estimate of the order of magnitude of Br emissions from volcanoes, we assume the inferred Br/S ratio to be representative of most volcanoes globally. Based on our Br/S ratio for all volcanic emissions, the estimated global annual sulphur emission of $14 \pm 6 \text{ Tg yr}^{-1}$ (ref. 19) would correspond to a source strength of the order of $30,000 \text{ t Br yr}^{-1}$, in agreement with literature estimates²⁰. The Cl/Br ratio for Indonesian²¹ and Japanese²² volcanoes has been found to be between 32 and 700. Taking an annual Cl emission by arc volcanoes (Montserrat belongs to this important type) of 3 Tg Cl yr^{-1} (ref. 23), our estimate of $0.03 \text{ Tg Br yr}^{-1}$ translates to a Cl/Br ratio of 100, well within the above range. Taking this range of Cl/Br ratios, the global volcanic BrO source strength would be 6,500 to $140,000 \text{ t yr}^{-1}$. The lower end of the range would still be quite important regionally. Should emissions be at the higher end of this range, then volcanoes would be the strongest single source of Br (exceeding CH₃Br with $\sim 100,000 \text{ t Br yr}^{-1}$)²⁴. In addition, bromine might also be present in the plume in other chemical forms (for example, HBr, HOBr), making the total Br emission even higher than the above figures indicate. Volcanoes are a potentially very important source of atmospheric bromine.

The present results show that volcanoes either directly emit reactive bromine (BrO) or emit bromine species that can be quickly converted to reactive bromine (that is, BrO) by gas-phase and/or heterogeneous reactions in the volcanic plume. The atmospheric lifetime of BrO_x can be very long (days to weeks)^{8,15}, because the

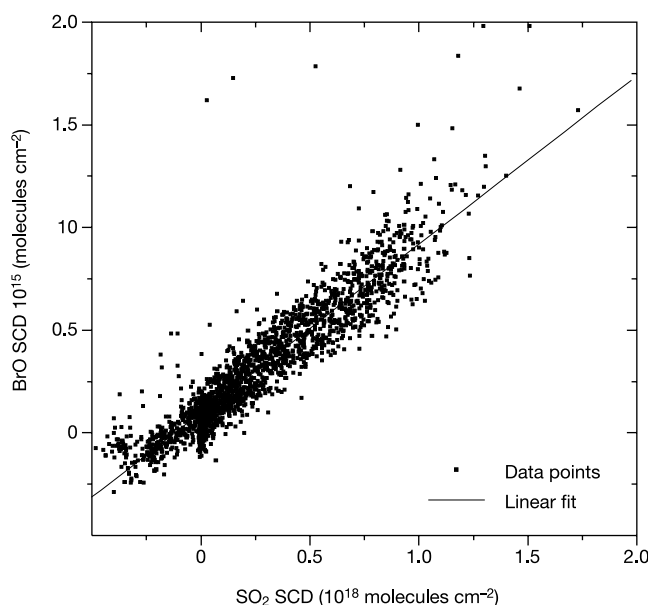


Figure 4 Plot of all data points, 23–31 May 2002. BrO is plotted as a function of SO₂; the equation of the linear fit is $\text{BrO} = 8.2 \times 10^{-4} \text{SO}_2 + 9.8 \times 10^{13}$; the species are well correlated with $R^2 = 0.81$. The molar ratio BrO/SO₂ is approximately 0.001.

recycling processes (converting HBr to BrO) are likely to be very efficient under the conditions in a volcanic plume. We conclude that local ozone depletion and small ozone 'holes' are likely to occur in the vicinity of active volcanoes, because 1 p.p.b. of BrO can destroy about 10 p.p.b. of ozone per minute. Ozone measurements close to the plume should be performed to investigate this effect. (We note that bromine-related total ozone losses in the polar troposphere and the Dead Sea basin have been observed at much lower BrO mixing ratios of ~30 parts per trillion (p.p.t.) and 180 p.p.t., respectively^{7–12}.) We also point out that the amounts of bromine emitted from volcanoes²⁵ are sufficiently large to play a role not only in the stratosphere²⁶, but also in tropospheric chemistry. Once released to the free troposphere, the lifetime of reactive Br is sufficiently long⁸ to allow regional and perhaps hemispheric mixing.

The relative abundance of BrO and SO₂ in the gas carries information about changes in the geophysical state of the volcano, which is of great importance for risk assessment. This information has been derived from HCl/SO₂ ratios determined by ground-based solar Fourier-transform infrared (FT-IR) spectroscopy²⁷. However, BrO/SO₂ measurements can be made simultaneously by optical spectrometers (described here), which are simpler than FT-IR spectrometers, are more easily deployed and automated, and allow a time resolution (~5 minutes) comparable to the time constants of variations in seismic signals. □

Methods

Data acquisition

The sequential scanning Mini-MAX-DOAS differs from the system described in ref. 11 in that a miniature spectrometer was used (OceanOptics Inc. USB2000, crossed Czerny-Turner arrangement, $1/f = 2.2$, 2,400 grooves mm⁻¹ grating, spectral range 251–402 nm, resolution 0.7 nm FWHM, CCD detector (2,048 elements at 12.5 μm centre-to-centre spacing) coupled to a 12-bit ADC connected to a PC via a USB interface). A Hoya U330 filter blocked the visible light at wavelengths >400 nm to reduce the stray light. The telescope was attached to a stepper motor mounted on a tripod to allow pointing it at angles between 10° and 90° above the horizon under computer control. In order to reduce the CCD dark current and temperature drift of the electronic offset signal and to stabilize the optical bench, the complete USB2000 spectrograph was kept at a temperature of 8 °C by Peltier cooling. To avoid water condensation, the whole unit was sealed in an argon-filled metal Dewar vessel; silica gel was added to keep the interior dry in case of leakage. The entire system (notebook PC, cooling system, spectrometer and stepper motor) operates on 12 V for about 24 hours from a car battery.

Data analysis

The SCDs of BrO and SO₂ were derived from the recorded spectra by the DOAS algorithm implemented in the WinDoas V2.10 software package from IASB (Belgium Institute for Space Aeronomy)²⁸. In the case of scattered sunlight as light source, the solar Fraunhofer lines (modulating $I_0(\lambda)$) have to be removed carefully in order to allow sensitive measurements of trace species. A Fraunhofer reference spectrum (FRS, $I_0(\lambda)$) was recorded each day, and care was taken that the FRS did not contain absorption by the volcanic plume. For the BrO evaluation, a spectral range containing eight absorption bands of the A²Π_{3/2} ← X²Π_{3/2} transition of the BrO molecule (320–360 nm, 542 spectral channels) was chosen. Literature reference spectra of BrO, NO₂, O₃, SO₂ and O₄ (all degraded to match the spectral resolution of our instrument by convolution with the instrumental function), a calculated 'Ring spectrum' (to compensate for reduction of the observed optical densities of solar Fraunhofer lines on long atmospheric light paths, that is, large solar zenith angles, due to rotational Raman scattering²⁹), a polynomial of fifth order (to remove broadband structures and the effects of Rayleigh and Mie scattering), and the FRS were simultaneously fitted to the measurement spectra using a nonlinear least-squares method³⁰. Sulphur dioxide was determined in the range from 307.5 to 316 nm (115 channels) encompassing four SO₂ absorption bands. Besides SO₂, reference spectra of NO₂, O₃, ClO, BrO, FRS and the 'Ring spectrum' as well as a second-order polynomial were included in the fit.

Received 13 November 2002; accepted 1 April 2003; doi:10.1038/nature01625.

- Francis, P., Burton, M. R. & Oppenheimer, C. Remote measurements of volcanic gas composition by solar occultation spectroscopy. *Nature* **396**, 567–570 (1998).
- McCormick, M. P., Thomason, L. W. & Trepte, C. R. Atmospheric effects of the Mt Pinatube eruption. *Nature* **373**, 399–404 (1995).
- Oppenheimer, C., Francis, P., Burton, M., Maciejewski, A. J. H. & Boardman, L. Remote measurement of volcanic gases by Fourier transform infrared spectroscopy. *Appl. Phys. B* **67**, 505–515 (1998).
- Pinto, J. P., Turco, R. P. & Toon, O. B. Self-limiting physical and chemical effects in volcanic eruption clouds. *J. Geophys. Res.* **94**, 11165–11174 (1989).
- Molina, M. J. & Rowland, F. S. Stratospheric sink for chlorofluoromethanes: Chlorine atom catalysed destruction of ozone. *Nature* **249**, 810–812 (1974).
- McElroy, M. B., Salawitch, R. J., Wofsy, C. S. & Logan, J. A. Reductions of Antarctic ozone due to synergistic interactions of chlorine and bromine. *Nature* **321**, 759–762 (1986).

- Hausmann, M. & Platt, U. Spectroscopic measurements of bromine oxide and ozone in the high Arctic during polar sunrise experiments 1992. *J. Geophys. Res.* **99**, 25399–25413 (1994).
- Barrie, L. & Platt, U. Arctic tropospheric chemistry: an overview. *Tellus B* **49**, 450–454 (1997).
- Tuckermann, M. *et al.* DOAS-observation of halogen radical-catalysed Arctic boundary layer ozone destruction during the ARCTOC campaign 1995 and 1996 in Ny-Alesund, Spitzbergen. *Tellus B* **49**, 533–555 (1997).
- Wagner, T. & Platt, U. Satellite mapping of enhanced BrO concentration in the troposphere. *Nature* **395**, 486–490 (1998).
- Hönninger, G. & Platt, U. Observations of BrO and its vertical distribution during surface ozone depletion at Alert. *Atmos. Environ.* **36**, 2481–2489 (2002).
- Hebestreit, K. *et al.* DOAS measurements of tropospheric bromine oxide of mid-latitudes. *Science* **283**, 55–57 (1999).
- Millan, M., Townsend, S. & Davies, J. Study of the Barringer refractor plate correlation spectrometer as a remote sensing instrument. Thesis, Univ. Toronto (1969).
- Galle, B. *et al.* A miniaturised ultraviolet spectrometer for remote sensing of SO₂ fluxes: a new tool for volcano surveillance. *J. Volcanol. Geotherm. Res.* **119**, 241–254 (2003).
- Platt, U. in *Monitoring by Spectroscopic Techniques* (ed. Sigrist, M. W.) 27–84 (Wiley & Sons, New York, 1994).
- Leser, H., Hönninger, G. & Platt, U. Max-Doas measurements of BrO and NO₂ in the marine boundary layer. *Geophys. Res. Lett.* doi:10.1029/2002GL015811 (2003).
- Young, S. R. *et al.* Overview of the eruption of Soufriere Hills Volcano, Montserrat, July 18, 1995, to December 1997. *Geophys. Res. Lett.* **25**, 3389–3392 (1998).
- Kraus, S. DOASIS, DOAS Windows software presentation. *Proc. 1st Int. DOAS-Workshop* (Heidelberg, Germany, 13–14 September 2001).
- Graf, H.-F., Feichter, J. & Langmann, V. Volcanic sulfur emissions: Estimates of source strength and its contribution to the global sulfate distribution. *J. Geophys. Res.* **102**, 10727–10738 (1997).
- Cadle, R. D. A comparison of volcanic with other fluxes of atmospheric trace gas constituents. *Rev. Geophys. Space Phys.* **18**, 4, 746–752 (1980).
- Kraft, K. & Chaigneau, M. Chlorine-bromine ratio of volcanic gas from twelve Indonesian volcanoes (Sumatra, Java, Flores, and Sulawesi). *C.R. Acad. Sci. D* **282**(4), 341–343 (1976).
- Sugiura, T., Mizutani, Y. & Oana, S. Fluorine, chlorine, bromine and iodine in volcanic gases. *J. Earth Sci. Nagoya Univ.* **11**, 272–278 (1963).
- Anderson, A. T. Some basaltic and andesitic gases. *Rev. Geophys. Space Phys.* **13**, 37–55 (1975).
- Butler, J. H. & Rodriguez, J. M. in *The Methyl Bromide Issue* (eds Bell, C., Price, N. & Chakrabarti, B.) 28–90 (Wiley, New York, 1996).
- Bluth, G. J. S., Doiron, S. D., Schnetzler, C. C., Krueger, A. J. & Walter, L. S. Global tracking of the SO₂ clouds from the June, 1991 Mount Pinatube eruptions. *Geophys. Res. Lett.* **19**, 151–154 (1992).
- Bureau, H., Kepler, H. & Metrich, N. Volcanic degassing of bromine and iodine: experimental fluid/melt partitioning data and applications to stratospheric chemistry. *Earth Planet. Sci. Lett.* **183**, 51–60 (2000).
- Francis, P., Maciejewski, A., Oppenheimer, C., Chaffin, C. & Caltabiano, T. SO₂:HCl ratios in the plume from Mt Etna and Vulcano determined by Fourier transform spectroscopy. *Geophys. Res. Lett.* **22**, 1717–1720 (1995).
- Van Roozendaal, M. & Fayt, C. *WinDOAS 2.1 Software User Manual* (BIRA-IASB, Uccle, 2001).
- Fish, D. J. & Jones, R. L. Rotational Raman scattering and the Ring effect in zenith-sky spectra. *Geophys. Res. Lett.* **22**, 811–814 (1995).
- Stutz, J. & Platt, U. Numerical analysis and estimation of the statistical error of differential optical absorption spectroscopy measurements with least squares methods. *Appl. Opt.* **35**(30), 6041–6053 (1996).

Acknowledgements We thank the staff of Montserrat Volcano Observatory, in particular G. Thomson, P. Dunkley and M. Edmonds, for their support during the field measurements.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to N.B. (Nicole.Bobrowski@iup.uni-heidelberg.de).

Parallel extinction risk and global distribution of languages and species

William J. Sutherland

Centre for Ecology, Evolution and Conservation, School of Biological Sciences, University of East Anglia, Norwich NR4 7TJ, UK

There are global threats to biodiversity with current extinction rates well above background levels¹. Although less well publicized, numerous human languages have also become extinct, and others are threatened with extinction^{2,3}. However, estimates of the number of threatened languages vary considerably owing to the wide range of criteria used. For example, languages have been classified as threatened if the number of speakers is less than 100,

Table 1 Comparisons of the threats to birds and languages in relation to population size and decline

Population size	Languages		Birds	
	Not declining	Declining	Not declining	Declining
Extinct	306 (4.4%)		125 (1.3%)	
1–49	256 (4.1%) CR	101 (1.6%) CR	39 (0.40%) CR	19 (0.20%) CR
50–249	379 (6.1%) EN	81 (1.3%) CR	30 (0.31%) EN	59 (0.61%) CR
250–999	691 (11.2%) VU	77 (1.2%) EN	67 (0.69%) VU	97 (1.0%) EN
1,000–2,499	634 (10.3%)	50 (0.8%) EN	?	103 (1.1%) EN
2,500–9,999	1,064 (17.2%)	41 (0.7%) VU	?	438 (4.5%) VU

Threat status (CR, critically endangered; EN, endangered; VU, vulnerable) is shown and was defined by using standard biological classification⁴. The proportion extinct is in relation to the number currently extant; the proportion threatened is in relation to the number assessed (that is, excluding deficient data).

500, 1,000, 10,000, 20,000 or 100,000 (ref. 3). Here I show, by applying internationally agreed criteria for classifying species extinction risk⁴, that languages are more threatened than birds or mammals. Rare languages are more likely to show evidence of decline than commoner ones. Areas with high language diversity also have high bird and mammal diversity and all three show similar relationships to area, latitude, area of forest and, for languages and birds, maximum altitude. The time of human settlement has little effect on current language diversity. Although similar factors explain the diversity of languages and biodiversity, the factors explaining extinction risk for birds and mammals (high altitude, high human densities and insularity) do not explain the numbers of endangered languages.

Species extinction risk is conventionally assessed by using standard quantitative criteria⁴ based on population size, actual or suspected population decline, range size changes and habitat fragmentation. Table 1 shows how these standard biological definitions⁴ of ‘critically endangered’, ‘endangered’ and ‘vulnerable’ relate to whether a population is declining and to population size. Species with small populations sometimes persist for long periods but have a considerably enhanced risk of extinction⁵ as a result of demographic stochasticity, environmental stochasticity, the Allee effect⁶ or genetic stochasticity⁷.

To compare the threats to biodiversity with those to languages, I used data⁸ on the estimated number of mother-tongue speakers along with any information on whether population declines have occurred to classify the extinction risk for all 6,809 living languages in the world, with the standard biological definitions (Table 1). I also counted the number of extinct languages, where documented, excluding any that went extinct before AD 1600 (the date conventionally used in species extinctions). For comparison I show the data⁹ on 9,797 species of birds analysed in exactly the same manner, except that they also include inferred declines, such as those due to habitat deterioration. Each data set is likely to have similar problems: the population sizes are estimates, some data are out of date, and some species and languages have undoubtedly gone extinct since the data were compiled (46 of the languages were classified as having just one speaker). This comparison suggests that the risks to languages greatly exceed those to birds: there are many more recorded language extinctions and substantially more rare languages. For example, 357 languages have fewer than 50 speakers.

My extinction risk classification for languages is conservative. The language database⁸ is excellent but lacks systematic data on range size, on whether quantified rapid declines have occurred, on

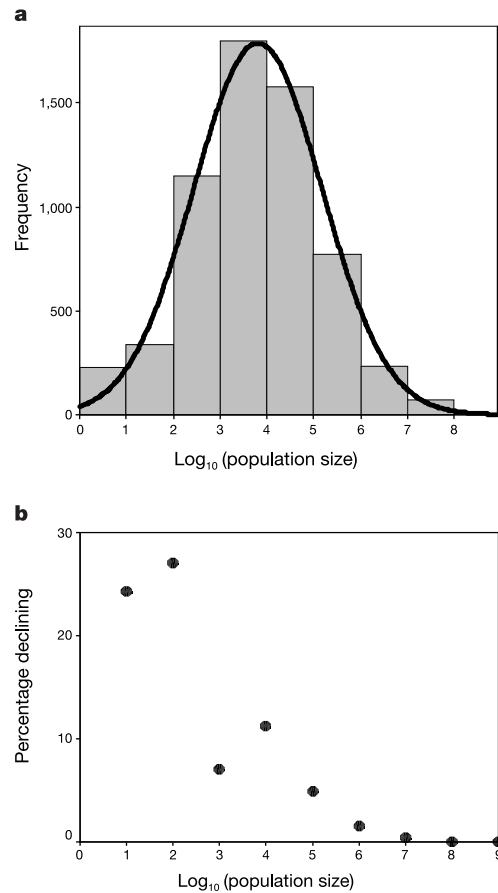


Figure 1 Language abundance and changes. **a**, Frequency distribution of extant global languages on a log₁₀ scale with a fitted normal distribution (mean 3.80, s.d. 1.38). **b**, Percentage of languages that are documented as declining in relation to the number of speakers.

whether the range is fragmented and on likely future declines—all of which are used by biologists to categorize extinction threats⁴. Of the birds classified as critically endangered on the basis of the entire data set⁹, only 64% (117 of 182) would have been classified as such on the basis of just population size and whether population decline is occurring. The equivalent figures are 58% (186 of 321) of endangered and 74% (505 of 680) of vulnerable birds. In the absence of more extensive information the threat status of many languages are similarly likely to be underestimated.

Table 2 compares the threat status of birds and mammals (based on all criteria: population size, decline, rapid decline, range size, habitat fragmentation and inferred threats) and languages, using the analysis from Table 1 (based just on population size and whether population decline is occurring). As described in the previous paragraph the analysis in Tables 1 and 2 underestimates the number of endangered languages. Even with this conservative comparison it is clear that the risks to languages exceed those to birds and mammals. Table 2 also shows that languages are at greater risk, both as a proportion of the total number and in absolute numbers.

Table 2 Comparisons of the threats to languages, birds and mammals: frequency distribution of languages, birds and mammals in each threat category

Category	Extinct	Critical	Endangered	Vulnerable	Data deficient	Total described extant
Languages	306 (4.5%)	438 (7.1%)	506 (8.2%)	732 (11.9%)	639	6,809
Birds	125 (1.3%)	182 (1.9%)	321 (3.3%)	680 (7.0%)	79	9,797
Mammals	87 (1.9%)	180 (4.1%)	340 (7.7%)	610 (13.9%)	240	4,630

Threat status of birds and mammals was assessed by using the complete IUCN criteria⁴, whereas the classification of languages is based solely on population size and decline, and is therefore underestimated. The proportion extinct is in relation to the number currently extant; the proportion threatened is in relation to the number assessed (that is, excluding deficient data).

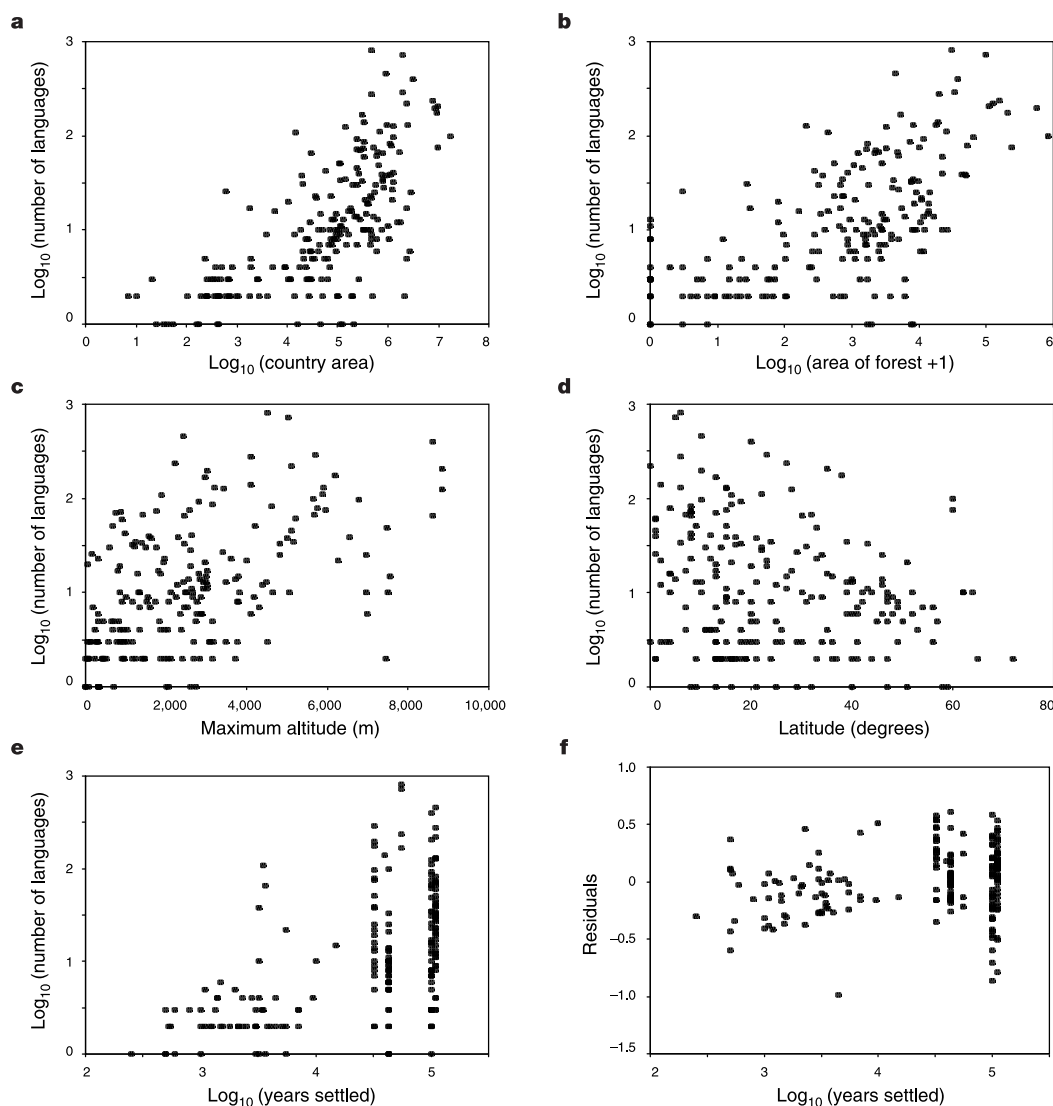


Figure 2 Factors correlating with language diversity. Relationships between $\log_{10}(\text{number of languages})$ and $\log_{10}(\text{area})$ ($r = 0.69$, $n = 214$, $P < 0.001$) (a), $\log_{10}(\text{area of forest} + 1)$ ($r = 0.63$, $n = 199$, $P < 0.001$) (b), maximum altitude ($r = 0.52$, $n = 214$, $P < 0.001$) (c), latitude ($r = -0.24$, $n = 214$, $P < 0.001$) (d), $\log_{10}(\text{number of years since first settlement})$ ($r = 0.48$, $n = 210$, $P < 0.001$) (e) and unstandardized residuals of the area relationship shown in a against $\log_{10}(\text{years settled})$ ($r = 0.14$, $n = 209$, $P < 0.041$) (f).

The frequency of languages of a given population size approximates to a log normal distribution (Fig. 1) with a small number of very abundant languages—a median of 6,300 mother-tongue speakers per language—although it is actually significantly different from normality (Kolmogorov–Smirnov test with Lilliefors significance correction = 0.24, d.f. = 6,168, $P < 0.001$) with a detrended Q–Q plot showing more very rare languages than expected.

Is there any evidence that languages might have a minimum viable size similar to the minimum population sizes of animal and plant populations? Figure 1b shows that the proportion of the languages that are documented as declining or predominantly elderly is negatively correlated with population size ($r_s = -0.966$, $n = 9$, $P < 0.001$). As with the Allee effect^{5,6}, this could mean that rarer languages will become even rarer and so go extinct. An obvious mechanism for such an Allee effect is that as languages become rare they become less attractive for people to learn and use.

There has been considerable interest in identifying areas of greatest biodiversity^{10,11}, and various factors affect global patterns of language diversity^{12–15}. I therefore sought to examine whether global linguistic

diversity and biodiversity are correlated and explained by the same variables. Countries with high language diversity also have high bird diversity ($r = 0.75$, $n = 204$, $P < 0.001$, both $\log_{10}$ transformed) and mammal diversity ($r = 0.69$, $n = 197$, $P < 0.001$, both $\log_{10}$ transformed). This analysis is obviously confounded by area, but the correlations remain once the effect of area has been removed: the residuals from the relationships between $\log_{10}(\text{number of languages or species})$ and $\log_{10}(\text{area})$ are correlated for both languages and birds ($r_s = 0.41$, $n = 204$, $P < 0.001$) and languages and mammals ($r_s = 0.41$, $n = 197$, $P < 0.001$).

The number of languages per country increases with country area (Fig. 2a), area of forest (Fig. 2b) and maximum altitude (Fig. 2c) but decreases with increasing latitude (Fig. 2d). Table 3 shows the results of a general linear model of the factors influencing numbers of languages and the number of bird and mammal species per country. Once the area effect has been incorporated, languages, birds and mammals are all more diverse in low-latitude countries, in countries with large areas of forest and, for birds and languages, in mountainous countries. Island countries have significantly fewer

Table 3 Correlates of the abundance of languages, birds and mammals

Variable	Languages		Birds		Mammals	
	F (sign)	P	F (sign)	P	F (sign)	P
Log ₁₀ (area)	39.09 (+)	0.001	33.45 (+)	0.001	20.61 (+)	0.001
Latitude	77.46 (–)	0.001	60.20 (–)	0.001	37.62 (–)	0.001
Log ₁₀ (forest area)	16.12 (+)	0.001	82.73 (+)	0.001	26.62 (+)	0.001
Altitude	4.12 (+)	0.044	4.68 (+)	0.032	0.40 (+)	0.530
Insularity	0.00 (+)	0.957	54.35 (–)	0.001	74.12 (–)	0.001
R ²	0.63		0.81		0.72	

GDP per person, growing season and date of modern human settlement were all considered in the full model but were not significant.

Table 4 Correlates of threatened languages, birds and mammals

Variable	Languages		Birds		Mammals	
	F (sign)	P	F (sign)	P	F (sign)	P
Total diversity	24.53 (+)	0.001	28.38 (+)	0.001	79.37 (+)	0.001
Log ₁₀ (area)	5.09 (+)	0.026	38.99 (+)	0.001	47.88 (+)	0.001
Growing season	8.76 (+)	0.004	14.93 (+)	0.001	1.10 (+)	0.296
Human density	0.04 (–)	0.851	13.83 (+)	0.001	4.57 (+)	0.034
Maximum altitude	0.03 (+)	0.868	19.85 (+)	0.001	7.49 (+)	0.007
Insularity	0.56 (–)	0.454	29.83 (–)	0.001	18.71 (–)	0.001
R ²	0.38		0.73		0.76	

Latitude, GDP per person, and log₁₀ forest area, were all considered but were not significant in the model for the number of languages. The numbers of endangered languages and species were square-root transformed prior to analysis.

birds and mammals but similar language diversity to non-island countries.

It has been suggested¹² that the length of the growing season (defined as the number of months in which the temperature exceeds 6 °C and rainfall in millimetres twice exceeds the average temperature in degrees Celsius¹²) influences language diversity but there was no evidence for this in this analysis. Furthermore, neither Gross Domestic Product (GDP) per person nor number of televisions per 1,000 people (as an index of national and global communication) was significantly related to the number of languages in a country.

Just as colonization and speciation explain increases in species within a new site, colonization and language divergence should result in increases in the numbers of languages after a country is first occupied. To examine whether time since settlement is currently important in determining language diversity I used the best estimates from refs 16 and 17 for the settlement of continents derived from archaeological and genetic data (Asia 100,000, Australia 55,000, Europe 43,000 and America 32,000 years BP) and a wide range of archaeological and historical information for settlement dates for islands. The time of settlement of the Americas is particularly controversial with current estimates ranging from 50,000 to 14,000 years BP. Language is thought to have appeared in Africa at some time between 50,000 and 150,000 years ago^{16,17}. It is generally accepted that the origin of languages predated the global spread of *Homo sapiens sapiens* and I used a date of 110,000 years BP.

Although the number of languages per country is significantly correlated with the number of years since settlement (Fig. 2e), many of the recently settled countries with few languages are small islands, and once the effect of area is removed, the relationship is greatly weakened (Fig. 2f). Settlement date was not significant when added to the general linear model (Table 3). Thus, the period since settlement has surprisingly little effect on language diversity. For example, the Pacific country of Vanuatu has 110 languages, yet the archaeological evidence shows only about 3,500 years of occupation. The two most likely reasons for duration's being relatively unimportant on a global scale are that languages can evolve so quickly that differences in settlement times are unimportant, and that processes such as the widespread loss of languages over much of Africa, Asia and Europe as a result of the development of Neolithic agriculture¹² suggest that languages are in a state of continual flux to which settlement time makes only a small contribution. This

supports the conclusion¹⁸, based on an analysis of linguistic stocks, that the high linguistic diversity within the Americas is not incompatible with the archaeological evidence that the Americas were occupied only relatively recently, such as 14,000 years BP.

Countries with the most endangered and extinct languages also have more endangered and extinct birds ($r_s = 0.36$, $n = 215$, $P < 0.001$) and mammals ($r_s = 0.34$, $n = 200$, $P < 0.001$). However, when expressed as a proportion of total numbers, the relationships between languages and species are not significant (birds: $r_s = -0.109$, $n = 215$, $P = 0.119$; mammals: $r_s = 0.03$, $n = 200$, $P = 0.662$). A general linear model analysis (Table 4) showed that, once the total number of species and area are controlled for, the numbers of endangered languages and birds are higher in areas with a long growing season (presumably because the lower ecological risk^{12,13} allows greater diversification). The numbers of endangered birds and mammals increase with human density (presumably because of greater habitat loss). There are more endangered birds and mammals in mountainous countries but fewer on islands. By contrast, the number of endangered languages did not vary with human density, altitude or insularity. It therefore seems that although patterns of language, bird and mammal diversity are similar, the reasons for extinction risk differ between cultural and biological diversity. □

Methods

Data sources were as follows: languages⁸, birds⁹, mammals¹⁹, total number of species¹⁹, closed non-plantation forest cover²⁰ in square kilometres (expressed as log₁₀(area of forest + 1), GDP²¹, climate²² for calculating growing season¹³, population altitude, area in square kilometres and latitude²³. Arithmetic means were taken if a range of values was given. Data for extinctions before 1600 AD were excluded. The bird data include three species classified as 'extinct in the wild', which I classified as 'extinct'. The data on bird extinctions²⁴ includes extinctions after 1500 AD. I excluded *Bulweria bifax*, *Pterodroma rupinarum*, *Arlantisia podarces*, *Porzana astricarpus*, *Dysmoropelia dekarchiskos* and *Upupa antaois* because these are all likely to have become extinct before 1600 AD.

Received 3 December 2002; accepted 21 March 2003; doi:10.1038/nature01607.

- Chapin, F. S. *et al.* Consequences of changing biodiversity. *Nature* **405**, 234–242 (2000).
- Krause, M. The world's languages in crisis. *Language* **68**, 4–10 (1992).
- Crystal, D. *Language Death* (Cambridge Univ. Press, Cambridge, 2000).
- IUCN/SSC. *Red List Categories and Criteria* (IUCN Species Survival Commission, Gland, 1994).
- Lande, R. Risks of population extinction from demographic and environmental stochasticity and random catastrophes. *Am. Nat.* **142**, 911–927 (1993).
- Stephens, P. A. & Sutherland, W. J. Consequences of the Allee effect for ecology and conservation. *Trends Ecol. Evol.* **14**, 401–405 (1999).
- Saccheri, I. *et al.* Inbreeding and extinction in a butterfly population. *Nature* **392**, 491–494 (1998).
- Grimes, B. F. (ed.) *Ethnologue: Languages of the World*, 14th edn (Summer Institute of Linguistics, Dallas, Texas, 2000).
- Birdlife International. *Threatened Birds of the World* (Lynx Edicions and Birdlife International, Barcelona and Cambridge, 2000).
- Balmford, A. *et al.* Conservation conflicts across Africa. *Science* **291**, 2616–2619 (2001).
- Myers, N. *et al.* Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- Nettle, D. *Linguistic Diversity* (Oxford Univ. Press, Oxford, 1999).
- Nettle, D. Explaining global patterns of language diversity. *J. Anthropol. Archaeol.* **17**, 354–374 (1998).
- Mace, R. & Pagel, M. A latitudinal gradient in the density of human languages in North America. *Proc. R. Soc. Lond. B* **261**, 117–121 (1995).
- Morre, J. L. *et al.* The distribution of cultural and biological diversity in Africa. *Proc. R. Soc. Lond. B* **269**, 1645–1653 (2002).
- Cavalli-Sforza, L. L., Menozzi, P. & Piazza, A. *The History and Geography of Human Genes* (Princeton Univ. Press, Princeton, New Jersey, 1994).
- Cavalli-Sforza, L. L. *Genes, Peoples and Languages* (Penguin, London, 2001).
- Nettle, D. Linguistic diversity of the Americas can be reconciled with a recent colonization. *Proc. Natl. Acad. Sci. USA* **96**, 3325 (1999).
- Groombridge, B. & Jenkins, M. D. *Global Biodiversity* (World Conservation Press, Cambridge, 2000).
- FAO. *State of the World's Forests 2001* (Food and Agriculture Organisation, Rome, 2000).
- World Factbook 2002 (<http://www.worldbank.org>).
- Mitchel, T., Hulme, M. & Newl, M. Climate data for political areas. *Area* **34**, 109–112 (2002).
- The Times Comprehensive Atlas of the World* (Times Books, London, 1999).
- Brooks, T. In *Threatened Birds of the World* (eds Birdlife International) 701–708 (Lynx Edicions and Birdlife International, Barcelona and Cambridge, 2000).

Acknowledgements I thank I. Côté, C. Crookford, R. Freckleton, J. Gill, P. Stephens and A. Watkinson for very useful comments.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to W.J.S. (w.sutherland@uea.ac.uk).

Rapid worldwide depletion of predatory fish communities

Ransom A. Myers & Boris Worm

Biology Department, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4J1

Serious concerns have been raised about the ecological effects of industrialized fishing^{1–3}, spurring a United Nations resolution on restoring fisheries and marine ecosystems to healthy levels⁴. However, a prerequisite for restoration is a general understanding of the composition and abundance of unexploited fish communities, relative to contemporary ones. We constructed trajectories of community biomass and composition of large predatory fishes in four continental shelf and nine oceanic systems, using all available data from the beginning of exploitation. Industrialized fisheries typically reduced community biomass by 80% within 15 years of exploitation. Compensatory increases in fast-growing species were observed, but often reversed within a decade. Using a meta-analytic approach, we estimate that large predatory fish biomass today is only about 10% of pre-industrial levels. We conclude that declines of large predators in coastal regions⁵ have extended throughout the global ocean, with potentially serious consequences for ecosystems^{5–7}. Our analysis suggests that management based on recent data alone may be misleading, and provides minimum estimates for unexploited communities, which could serve as the ‘missing baseline’⁸ needed for future restoration efforts.

Ecological communities on continental shelves and in the open ocean contribute almost half of the planet’s primary production⁹, and sustain three-quarters of global fishery yields¹. The widespread decline and collapse of major fish stocks has sparked concerns about the effects of overfishing on these communities. Historical data from coastal ecosystems suggest that losses of large predatory fishes,

as well as mammals and reptiles, were especially pronounced, and precipitated marked changes in coastal ecosystem structure and function⁵. Such baseline information is scarce for shelf and oceanic ecosystems. Although there is an understanding of the magnitude of the decline in single stocks¹⁰, it is an open question how entire communities have responded to large-scale exploitation. In this paper, we examine the trajectories of entire communities, and estimate global rates of decline for large predatory fishes in shelf and oceanic ecosystems.

We attempted to compile all data from which relative biomass at the beginning of industrialized exploitation could be reliably estimated. For shelf ecosystems, we used standardized research trawl surveys in the northwest Atlantic Ocean, the Gulf of Thailand and the Antarctic Ocean off South Georgia, which were designed to estimate the biomass of large demersal fish such as codfishes (Gadidae), flatfishes (Pleuronectidae), skates and rays (Rajidae), among others (see Supplementary Information for detailed species information). In all other shelf areas for which we could obtain data, industrialized trawl fisheries occurred before research surveys took place. For oceanic ecosystems, we used Japanese pelagic longlining data, which represent the complete catch-rate data for tuna (Thunnini), billfishes (Istiophoridae) and swordfish (Xiphiidae) aggregated in monthly intervals, from 1952 to 1999, across a global 5° × 5° grid. Pelagic longlines are the most widespread fishing gear, and the Japanese fleet the most widespread longline operation, covering all oceans except the circumpolar seas. Longlines, which resemble long, baited transects, catch a wide range of species in a consistent way and over vast spatial scales. We had to restrict our analysis of longlining data to the equatorial and southern oceans, because industrialized exploitation was already underway in much of the Northern Hemisphere before these data were recorded^{11,12}. Longlining data were separated into temperate, subtropical and tropical communities (see Methods).

For each shelf and oceanic community, i , we estimated

$$N_i(t) = N_i(0)[(1 - \delta_i)e^{-r_i t} + \delta_i] \quad (1)$$

where $N_i(t)$ is the biomass at time t , $N_i(0)$ is the initial biomass

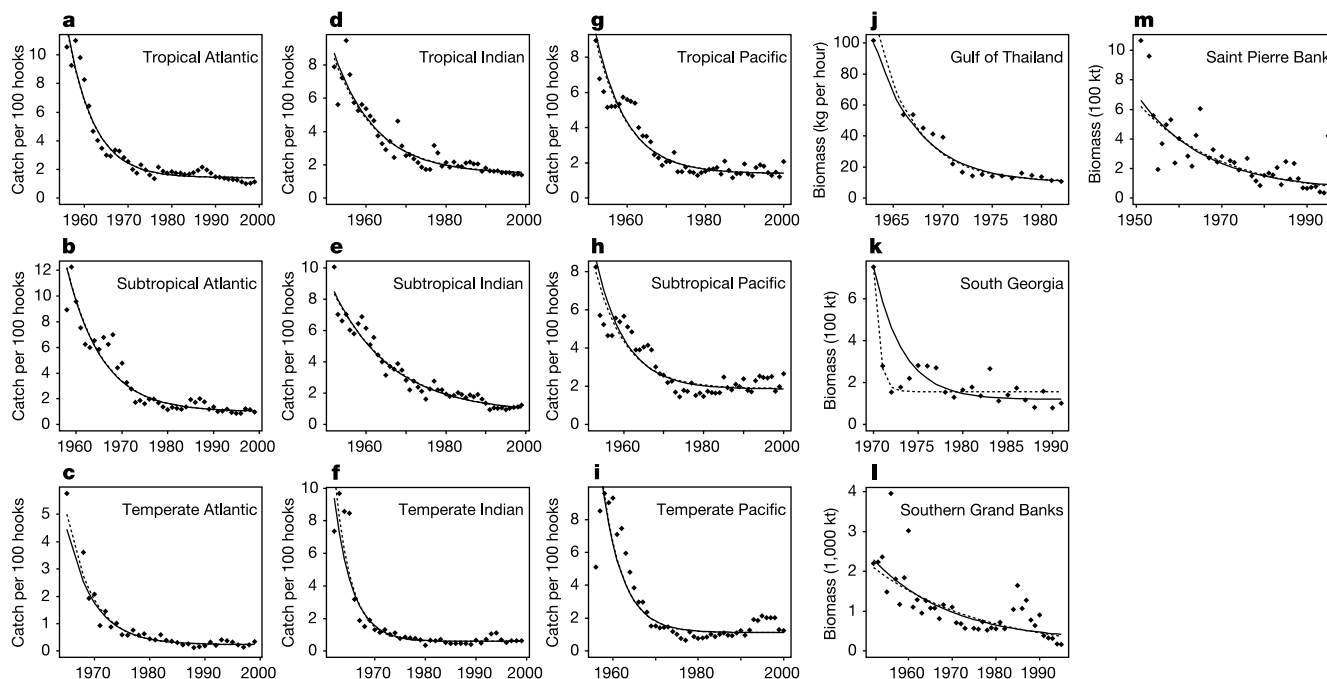


Figure 1 Time trends of community biomass in oceanic (a–i) and shelf (j–m) ecosystems. Relative biomass estimates from the beginning of industrialized fishing (solid

points) are shown with superimposed fitted curves from individual maximum-likelihood fits (solid lines) and empirical Bayes predictions from a mixed-model fit (dashed lines).

Table 1 Meta-analysis of time trends in predatory fish biomass

Region	$r_i (\times 100)$			$\delta_i (\times 100)$		
	Individual fit	CL	Mixed model	Individual fit	CL	Mixed model
Tropical Atlantic	16.6	13.5–19.7	16.7	12.1	10.0–14.5	11.9
Subtropical Atlantic	12.9	10.1–15.7	13.0	8.1	6.4–10.2	8.3
Temperate Atlantic	21.4	15.8–26.9	20.3	4.7	3.2–6.9	5.3
Tropical Indian	9.2	7.1–11.4	9.5	17.6	14.9–20.6	16.8
Subtropical Indian	6.5	5.1–7.8	6.8	8.2	5.5–12.3	9.2
Temperate Indian	30.7	23.7–37.8	27.7	5.5	3.9–7.7	6.3
Tropical Pacific	12.1	9.4–14.8	12.4	15.5	13.0–18.6	14.9
Subtropical Pacific	12.8	8.5–17.1	13.5	23.5	18.9–29.3	21.5
Temperate Pacific	20.8	14.3–27.3	20.4	8.2	5.6–12.1	8.5
Gulf of Thailand	25.6	18.2–33.0	22.2	9.3	6.8–12.6	9.8
South Georgia	166.6	2.2–331.1	30.8	20.9	17.5–25.0	16.0
Southern Grand Banks	4.0	2.9–5.1	5.7	0.0	–	10.0
Saint Pierre Banks	5.1	0.1–10.1	6.3	2.7	0.0–36600	7.9
Mixed model mean			16.0			10.3
Mixed model CL			10.7–21.3			7.7–13.9
Distribution			4.5–31.6			4.6–23.6

Two parameters were estimated: r_i is the initial rate of decline (in per cent per year), and δ_i the residual biomass proportion at equilibrium (in per cent). Point estimates and 95% confidence limits (CL) are presented for the individual maximum likelihood fits, and for the mixed-effects model that combined all data (see Methods for details). The random-effects distribution (95% limits) provides a measure of the estimated parameter variability across communities.

before industrialized exploitation, and r_i is the initial rate of decline to δ_i , the fraction of the community that remains at equilibrium. The initial rate of decline in total biomass—that is, the fraction lost in the first year—is $(1 - \delta_i)(1 - e^{-r_i})$. Then we combined all data using nonlinear mixed-effects models¹³, where $r_i \sim N(\mu_r, \sigma_r^2)$ and $\log \delta_i \sim N(\mu_\delta, \sigma_\delta^2)$, to estimate a global mean and variance of r_i and δ_i .

In the open ocean communities, we observed surprisingly consistent and rapid declines, with catch rates falling from 6–12 down to 0.5–2 individuals per 100 hooks during the first 10 years of exploitation (Fig. 1a–i). Rates of decline were similar in tropical and subtropical regions, but consistently highest in temperate regions in all three oceans (Fig. 1c, f, i and Table 1). Temperate regions also showed the lowest equilibrium catch rates (Table 1). Spatial pattern

of expansion and decline of pelagic fisheries are shown in Fig. 2. During the global expansion of longline fisheries in the 1950s to 1960s, high abundances of tuna and billfish were always found at the periphery of the fished area (Fig. 2a–c). Most newly fished areas showed very high catch rates, but declined to low levels after a few years. As a result, all areas now sustain low catch rates, and some formerly productive areas have been abandoned (Fig. 2d). In shelf communities, we observed declines of similar magnitude as in the open ocean. The Gulf of Thailand, for example, lost 60% of large finfish, sharks and skates during the first 5 years of industrialized trawl fishing (Fig. 1j). The highest initial rate of decline was seen in South Georgia (Fig. 1k), which has a narrow shelf area that was effectively fished down during the first 2 years of exploitation¹⁴. Less-than-average declines were seen on the Southern Grand Banks

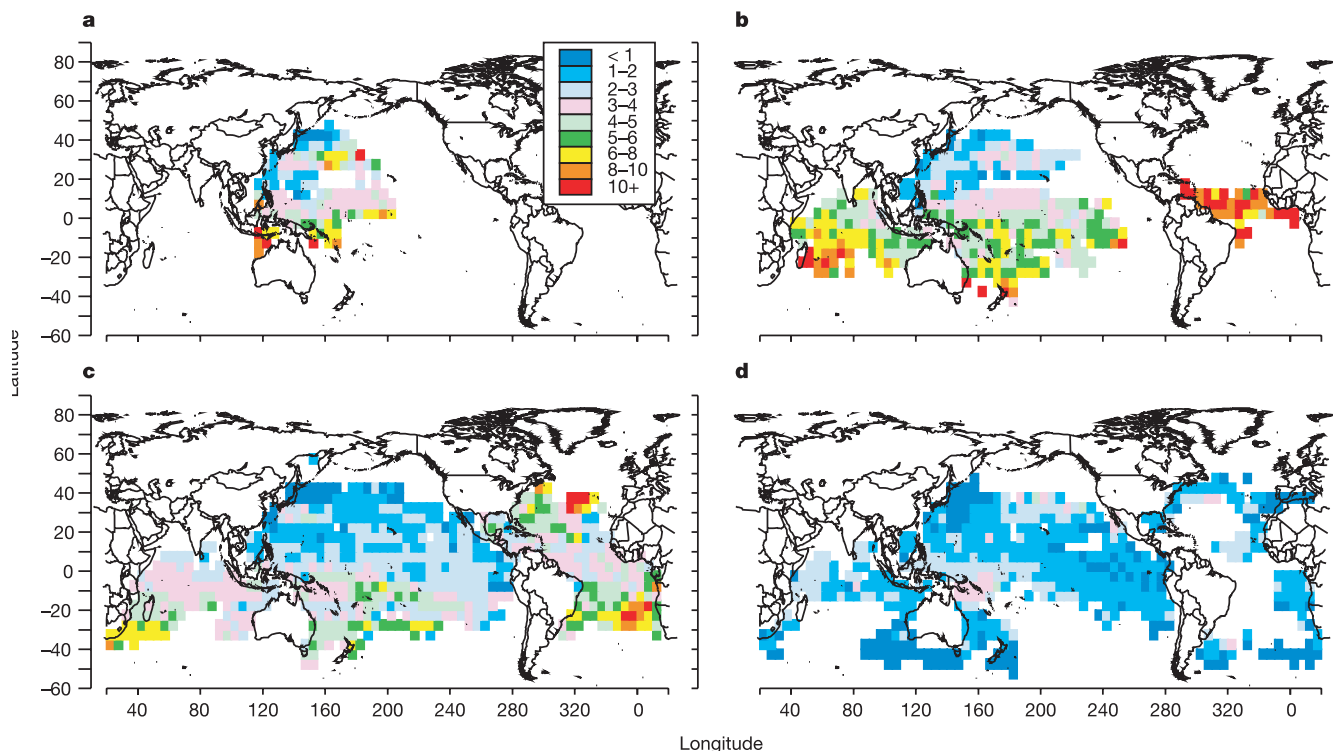


Figure 2 Spatial patterns of relative predator biomass in 1952 (a), 1958 (b), 1964 (c) and 1980 (d). Colour codes depict the number of fish caught per 100 hooks on pelagic

longlines set by the Japanese fleet. Data are binned in a global $5^\circ \times 5^\circ$ grid. For complete year-by-year maps, refer to the Supplementary Information.

(Fig. 1l) and Saint Pierre Bank (Fig. 1m); these communities may already have been affected by intense pre-industrial fisheries¹⁵.

By combining all data using a mixed-effects model, we estimated that the mean initial rate of decline, r_i , is 16% per year, and the mean residual equilibrium biomass, δ_i , is 10% of pre-exploitation levels (Table 1). So, an 80% decline typically occurred within 15 years of industrialized exploitation, which is usually before scientific monitoring has taken place. The proportion of residual biomass, δ_i , showed remarkably little variation between communities (Table 1): the mixed-effects model estimates imply that 95% of communities would have a residual biomass proportion between 5% and 24%. We believe that these still represent conservative estimates of total predator declines for the following reasons: (1) pre-industrial removals from some of the shelf communities¹⁵; (2) gear saturation at high catch rates in the early longlining data, as well as higher initial levels of shark damage leading to an underestimation of initial biomass¹⁶ (see Supplementary Information); (3) increasing fishing power of longline vessels over time owing to improved navigation and targeting of oceanographic features¹⁷; and (4) targeting of some migratory species, such as southern bluefin tuna (*Thunnus maccoyii*), at their tropical spawning grounds before widespread exploitation in temperate areas occurred¹⁸. Furthermore, declines in other large predators such as sharks are not fully captured by our data, but may be of similar or greater magnitude than those of bony fishes^{19,20}.

One mechanism that could compensate for the effects of overfishing is the increase in non-target species due to release from predation or competition²¹. In our analyses, we see evidence for species compensation in both oceanic billfish and shelf groundfish communities (Fig. 3). According to the longlining data and to early surveys^{11,12}, blue marlin was initially the dominant billfish species, but declined rapidly in the 1950s (Fig. 3a). Simultaneous increases in faster-growing species such as sailfish were observed, followed by a decrease, possibly due to increased 'bycatch' mortality (Fig. 3a; neither species was targeted by the Japanese fleet). Coincidentally, swordfish catch rates increased until these fish became prime targets of other fleets in the late 1980s. Surprisingly consistent patterns of compensatory increase and decline were seen in most pelagic communities (see Supplementary Information). Similarly, in the North Atlantic demersal communities, we observed rapid initial declines, particularly in large codfishes, but also in skates and rockfish. Although the dominant codfishes declined sixfold between 1952 and 1970, sixfold increases were seen in the flatfishes, which were not initially targeted by the trawl fishery (Fig. 3b). Some increase in the gadoids occurred when implementation of the 200-mile limit in 1977 curtailed foreign overfishing in Canadian waters. However, as in the billfish data, we observed an ultimate decline in all species groups (Fig. 3b) as fishing pressure from Canadian and other fleets intensified in the late 1980s, leading to the collapse of all major groundfish stocks¹⁰. We conclude that some species compensation was evident, but often reversed within a decade or less, probably because of changes in targeting or bycatch.

Our analysis suggests that the global ocean has lost more than 90% of large predatory fishes. Although it is now widely accepted that single populations can be fished to low levels, this is the first analysis to show general, pronounced declines of entire communities across widely varying ecosystems. Although the overall magnitude of change is evident, there remains uncertainty about trajectories of individual tuna and billfish species. Assessments of these species are continually improved by the international management agencies. However, most scientists and managers may not be aware of the true magnitude of change in marine ecosystems, because the majority of declines occurred during the first years of exploitation, typically before surveys were undertaken. Management schemes are usually implemented well after industrialized fishing has begun, and only serve to stabilize fish biomass at low levels. Supporting evidence for these conclusions comes from the United Nations Food and Agriculture Organization (FAO) data set, which indicates declining global catches²² and a consistent decline in the mean trophic level of the catch²³, which is a result of removing predatory fishes. Furthermore, on seamounts and on continental slopes, where virgin communities are fished, similar dynamics of extremely high catch rates are observed, which decline rapidly over the first 3–5 years of exploitation²⁴. We suggest that this pattern is not unique to these communities, but simply a universal feature of the early exploitation of ecosystems.

Our results have several important management implications. First, we need to consider potential ecosystem effects of removing 90% of large predators. Fishery-induced top-down effects are evident in coastal⁵ and shelf²⁵ ecosystems, but little empirical information is available from the open oceans. This is worrisome, as any ecosystem-wide effect is bound to be widespread, and possibly difficult to reverse, because of the global scale of the declines (Fig. 2). Another serious problem in heavily depleted communities is the extinction of populations, particularly those with high ages of maturity²⁶. Local extinctions can go unnoticed even in closely monitored systems such as the northwest Atlantic²⁷, let alone in the open ocean. Finally, the reduction of fish biomass to low levels may compromise the sustainability of fishing, and support only relatively low economic yields³. Such concerns have motivated a recent UN resolution to restore fish stocks to healthy

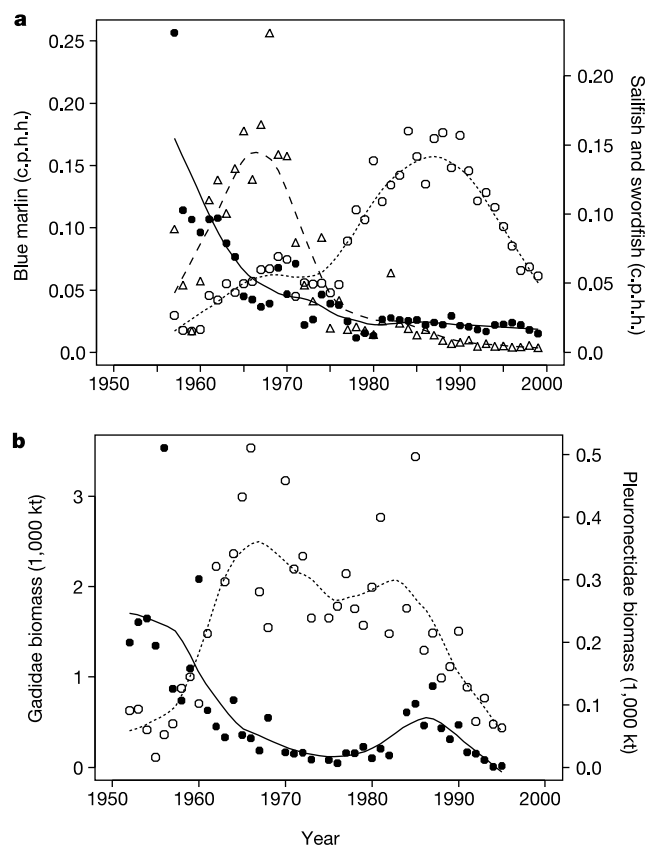


Figure 3 Compensation in exploited fish communities. **a**, Oceanic billfish community in the tropical Atlantic, showing the catch per 100 hooks (c.p.h.h.) of blue marlin (*Makaira nigricans*; solid circles, solid line), sailfish (*Istiophorus platypterus*; open triangles, dashed line) and swordfish (*Xiphias gladius*; open circles, dotted line). **b**, Demersal fish community on the Southern Grand Banks, showing the biomass of codfishes (Gadidae; solid circles, solid line) and flatfishes (Pleuronectidae; open circles, dotted line). Lines represent best fits using a local regression smoother.

levels⁴. Our analysis shows that it is appropriate and necessary to attempt restoration on a global scale, and provides a benchmark against which community recovery could be assessed. □

Methods

Data selection

For shelf communities, we compiled data from research trawl surveys from the Southern Grand Banks (43–46° N, 49–53° W) and Saint Pierre Banks (45–47° N, 55–58° W) (ref. 28), the Gulf of Thailand (9–14° N, 100–105° W) (ref. 29) and South Georgia (53–56° S, 35–40° W) (ref. 14). All other trawl data sets that we considered (for example, North Sea, Georges Bank and Alaska) did not capture the beginning of industrialized exploitation. We included only demersal predators; pelagic species, which were not well sampled by the trawl gear, were excluded. Longlining data obtained from the Japanese Fishery Agency were divided into temperate (Atlantic, 40–45° S; Indian, 35–45° S; Pacific, 30–45° S), subtropical (Atlantic, 10–40° S; Indian, 10–35° S; Pacific, 15–30° S) and tropical communities (Atlantic, 20° N–10° S; Indian, 15° N–15° S; Pacific, 10–15° S). These divisions were based on their dominant species: yellowfin (*T. albacares*), albacore (*T. alalunga*) or southern bluefin tuna (*T. maccoyii*), respectively, and excluded areas previously fished by the Japanese, Spanish and US fleets. Running the models with alternative divisions ($\pm 5^\circ$) did not change the results significantly. The catch rates in each community were determined as the sum of the catches divided by the sum of the effort in each region in each year. Years with very low effort (<20,000 hooks for the entire region) were excluded. Alternative treatment of the data, including removing seasonal effects and taking the average catch rates over $5^\circ \times 5^\circ$ squares, had little effect on the results. For longlines, we assume that the catch rate is an approximate measure of relative biomass, which is probably conservative because the average individual weights of fish in exploited populations tend to decline over time. Our data capture the abundance of larger fishes that are vulnerable to baited hooks and bottom trawls, respectively. Many smaller species have low catchabilities and are not recorded reliably by these methods. Changes in the longline fishery occurred around 1980 when the fishery began to expand into deeper regions; however, this was only after the declines in biomass were observed. For more details on species composition, data treatment and interpretation of trends, refer to the Supplementary Information.

Data analysis

Our model (equation (1)) assumes that for each community, i , the rate of decline to equilibrium is exponential with rate r_i from a pre-exploitation biomass $N_i(0)$, where $t = 0$ is the first year of industrialized fishing. Exploitation continues until equilibrium is approached, where a residual proportion, δ_i , of the biomass remains. We fit this model separately to each community under the assumption of a lognormal error distribution using nonlinear regression (Procedure NLIN in SAS, version 8). We also used nonlinear mixed-effects models¹³ to determine whether the patterns were similar across communities. Mixed-effect models were fitted by maximizing the likelihood integrated over the random effects using adaptive gaussian quadrature (Procedure NLMIXED in SAS). To account for the fact that biomass was recorded in different units (kilotonnes (kt), catch rates), the initial biomass, $N_i(0)$, was assumed to be a fixed effect for each community with appropriate units. For South Georgia, $N_i(0)$ was fixed at the first biomass estimate to capture the high initial rate of decline. This first estimate (750 kt; ref. 14) was considered to be realistic because it was very close to the sum of total removals (514 kt; ref. 30) plus the residual biomass estimate (160 kt; ref. 14) after the first 2 years of fishing. Autocorrelation in the residuals of some time series may cause the standard errors to be underestimated. The results were robust to alternative error assumptions (separate error variances for the time series and alternative error distributions); for example, under the assumption of normal errors, the rate of decline was 13.9% and residual biomass was 10.9%, respectively.

Received 25 November 2002; accepted 25 March 2003; doi:10.1038/nature01610.

1. Pauly, D. & Christensen, V. Primary production required to sustain global fisheries. *Nature* **374**, 255–257 (1995).
2. Tegner, M. J. & Dayton, P. K. Ecosystem effects of fishing. *Trends Ecol. Evol.* **14**, 261–262 (1999).
3. Pauly, D. *et al.* Towards sustainability in world fisheries. *Nature* **418**, 689–695 (2002).
4. United Nations. *World Summit on Sustainable Development: Plan of Implementation* (http://www.johannesburgsummit.org/html/documents/summit_docs/2309_planfinal.htm) (United Nations, New York, 2002).
5. Jackson, J. B. C. *et al.* Historical overfishing and the recent collapse of coastal ecosystems. *Science* **293**, 629–638 (2001).
6. Steele, J. H. & Schumacher, M. Ecosystem structure before fishing. *Fish. Res.* **44**, 201–205 (2000).
7. Worm, B., Lotze, H. K., Hillebrand, H. & Sommer, U. Consumer versus resource control of species diversity and ecosystem functioning. *Nature* **417**, 848–851 (2002).
8. Pauly, D. Anecdotes and the shifting baseline syndrome of fisheries. *Trends Ecol. Evol.* **10**, 430 (1995).
9. Field, C. B., Behrenfeld, M. J., Randerson, J. T. & Falkowski, P. Primary production of the biosphere: integrating terrestrial and oceanic components. *Science* **281**, 237–240 (1998).
10. Myers, R. A., Hutchings, J. A. & Barrowman, N. J. Why do fish stocks collapse? The example of cod in Atlantic Canada. *Ecol. Appl.* **7**, 91–106 (1997).
11. Bullis, H. R. Preliminary report on exploratory longline fishing for tuna in the Gulf of Mexico and the Caribbean Sea. *Comm. Fish. Rev.* **17**, 1–15 (1955).
12. Shomura, R. S. & Murphy, G. I. *Longline Fishery for Deep-Swimming Tunas in the Central Pacific, 1953*. US Fish and Wildlife Service Special Scientific Report: Fisheries No. 157 (USFWS, Washington, 1955).
13. Davidian, M. & Giltinan, D. M. *Nonlinear Models for Repeated Measurement Data* (Chapman & Hall, New York, 1995).

14. Kock, K.-H. & Shimadzu, Y. in *Southern Ocean Ecology: The BIOMASS Perspective* (ed. El-Sayed, S. Z.) 287–312 (Cambridge Univ. Press, Cambridge, 1994).
15. Hutchings, J. A. & Myers, R. A. in *The North Atlantic Fisheries: Successes, Failures and Challenges* (eds Arnason, R. & Felt, L.) 38–93 (The Institute of Island Studies, Charlottetown, Prince Edward Island, Canada, 1995).
16. Rothschild, B. J. Competition for gear in a multiple-species fishery. *J. Cons. Int. Explor. Mer.* **31**, 102–110 (1967).
17. Lyne, V., Parslow, J., Young, J., Pearce, A. & Lynch, M. *Development, Application and Evaluation of the Use of Remote Sensing Data by Australian Fisheries* (CSIRO Marine Research, Hobart, Australia, 2000).
18. Caton, A. E. A review of aspects of southern bluefin tuna biology, population and fisheries. *Inter-Am. Trop. Tuna Comm. Spec. Rep.* **7**, 181–350 (1991).
19. Stevens, J. D., Bonfil, R., Dulvy, N. K. & Walker, P. A. The effects of fishing on sharks, rays, and chimaeras (chondrichthyans), and the implications for marine ecosystems. *ICES J. Mar. Sci.* **57**, 476–494 (2000).
20. Baum, J. K. *et al.* Collapse and conservation of shark populations in the Northwest Atlantic. *Science* **299**, 389–392 (2003).
21. Fogarty, M. J. & Murawski, S. A. Large-scale disturbance and the structure of marine systems: Fishery impacts on Georges Bank. *Ecol. Appl.* **8** (suppl.), 6–22 (1998).
22. Watson, R. & Pauly, D. Systematic distortions in world fisheries catch trends. *Nature* **414**, 534–536 (2001).
23. Pauly, D., Christensen, V., Dalsgaard, J., Froese, R. & Torres, F. Jr Fishing down marine food webs. *Science* **279**, 860–863 (1998).
24. Roberts, C. M. Deep impact: the rising toll of fishing in the deep sea. *Trends Ecol. Evol.* **242**, 242–245 (2002).
25. Worm, B. & Myers, R. A. Meta-analysis of cod–shrimp interactions reveals top–down control in oceanic food webs. *Ecology* **84**, 162–173 (2003).
26. Myers, R. A. & Mertz, G. The limits of exploitation: a precautionary approach. *Ecol. Appl.* **8** (suppl.), 165–169 (1998).
27. Casey, J. M. & Myers, R. A. Near extinction of a large, widely distributed fish. *Science* **281**, 690–692 (1998).
28. Casey, J. M. *Fish Assemblages on the Grand Banks of Newfoundland*. Thesis, Memorial Univ. Newfoundland, Canada (2000).
29. Pauly, D. in *Fish Population Dynamics* (ed. Gulland, J. A.) 329–348 (Wiley, New York, 1988).
30. Commission for the Conservation of Antarctic Marine Living Resources *Statistical Bulletin Vol. 14* (CCAMLR, Hobart, Australia, 2002).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank J. Casey, A. Fonteneau, S. Hall, J. Hampton, S. Harley, J. Ianelli, I. Jonsen, J. Kitchell, K.-H. Kock, H. Lotze, M. Maunder, T. Nishida, M. Prager, T. Quinn, G. Scott and P. Ward for data, comments and suggestions, N. Barrowman and W. Blanchard for statistical advice, and D. Swan for technical assistance. This research is part of a larger project on pelagic longlining initiated and supported by the Pew Charitable Trusts. Further support was provided by the Deutsche Forschungsgemeinschaft and the National Sciences and Engineering Research Council of Canada.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.A.M. (Ransom.Myers@dal.ca).

Attractor dynamics of network UP states in the neocortex

Rosa Cossart, Dmitriy Aronov & Rafael Yuste

Department of Biological Sciences, Columbia University, New York, New York 10027, USA

The cerebral cortex receives input from lower brain regions, and its function is traditionally considered to be processing that input through successive stages to reach an appropriate output^{1,2}. However, the cortical circuit contains many interconnections, including those feeding back from higher centres^{3–6}, and is continuously active even in the absence of sensory inputs^{7–9}. Such spontaneous firing has a structure that reflects the coordinated activity of specific groups of neurons^{10–12}. Moreover, the membrane potential of cortical neurons fluctuates spontaneously between a resting (DOWN) and a depolarized (UP) state^{11,13–16}, which may also be coordinated. The elevated firing rate in the UP state follows sensory stimulation¹⁶ and provides a substrate for

persistent activity, a network state that might mediate working memory^{17–21}. Using two-photon calcium imaging, we reconstructed the dynamics of spontaneous activity of up to 1,400 neurons in slices of mouse visual cortex. Here we report the occurrence of synchronized UP state transitions ('cortical flashes') that occur in spatially organized ensembles involving small numbers of neurons. Because of their stereotyped spatiotemporal dynamics, we conclude that network UP states are circuit attractors—emergent features of feedback neural networks²² that could implement memory states or solutions to computational problems.

To investigate the spatiotemporal dynamics and mechanisms of persistent neocortical activity, we used two-photon calcium imaging to reconstruct the spontaneous activity of large neuronal populations with single-cell resolution (184–1,396 cells, 650 on average; Fig. 1). Unstimulated neocortical slices perfused with standard artificial cerebrospinal fluid (aCSF) always exhibited persistent spontaneous activity, as detected by imaging spontaneous calcium transients¹² (Fig. 1c). All calcium transients were blocked by the bath application of the sodium channel blocker tetrodotoxin (1 μ M, $n = 14$; not shown) or voltage-sensitive calcium channel (VSCC) antagonists (2 mM Ni^{2+} or 200 μ M Cd^{2+} , $n = 13$; not shown). As in our previous work¹², we concluded that calcium transients were due to the opening of somatic VSCCs by sodium action potentials, and that they could be used as an indicator that action potentials had occurred in a given cell.

To analyse the spontaneous activity of the entire network, we measured fluorescence changes in each cell and recorded the peak time of each calcium transient (Fig. 1c; see Methods). We then combined calcium transients from all cells into rasterplots and

collapsed these rasterplots into activity histograms, which indicated the percentage of active cells as a function of time (Fig. 1d). On average, $0.4 \pm 0.1\%$ of cells were active within any given 1-s interval ($n = 168$ movies). In 78% of movies, these time histograms showed prominent peaks of synchronous activity (Fig. 1d, asterisks). Cells that were coactive during peaks of synchrony constituted a low, but significant, fraction of all imaged cells ($2.2 \pm 0.1\%$, $n = 414$). This is a considerably smaller fraction of cells than that active during other synchronous network phenomena in slices, such as epileptiform²³ or oscillatory¹¹ events. Furthermore, forms of persistent activity previously reported *in vitro* in modified slice media, or *in vivo*, were paroxysmal and propagative events that recruit most neurons^{10,11,13–15}. Also, peaks of synchrony had no apparent periodicity and occurred with an average interval of 55 ± 4 s ($n = 249$ intervals), considerably less often than both epileptiform events (~ 1 Hz) and slow oscillations (0.1–0.5 Hz). Thus, the synchronizations we describe represent a different type of spontaneous event.

Synchronized neuronal ensembles were often spatially structured (Fig. 2; see Methods). At peaks of synchrony, coactive cells were either distributed across the imaged field (74% of peaks) or organized into clusters (19%), layers (4%) or columns (3%). Clusters were found in layers II/III or V, whereas layers and columns could involve any layer. Interestingly, many peaks of synchronous activity were preceded by a gradual build-up and followed by a gradual decrease in the number of active cells. In some cases, a significant fraction of cells was active up to 5 s before and after the peak of synchrony (Fig. 3A). In addition, the spatial arrangement of coactive cells gradually became organized during the period of time preceding the peak (Fig. 3B). Even 5 s away from the peaks, 13% of

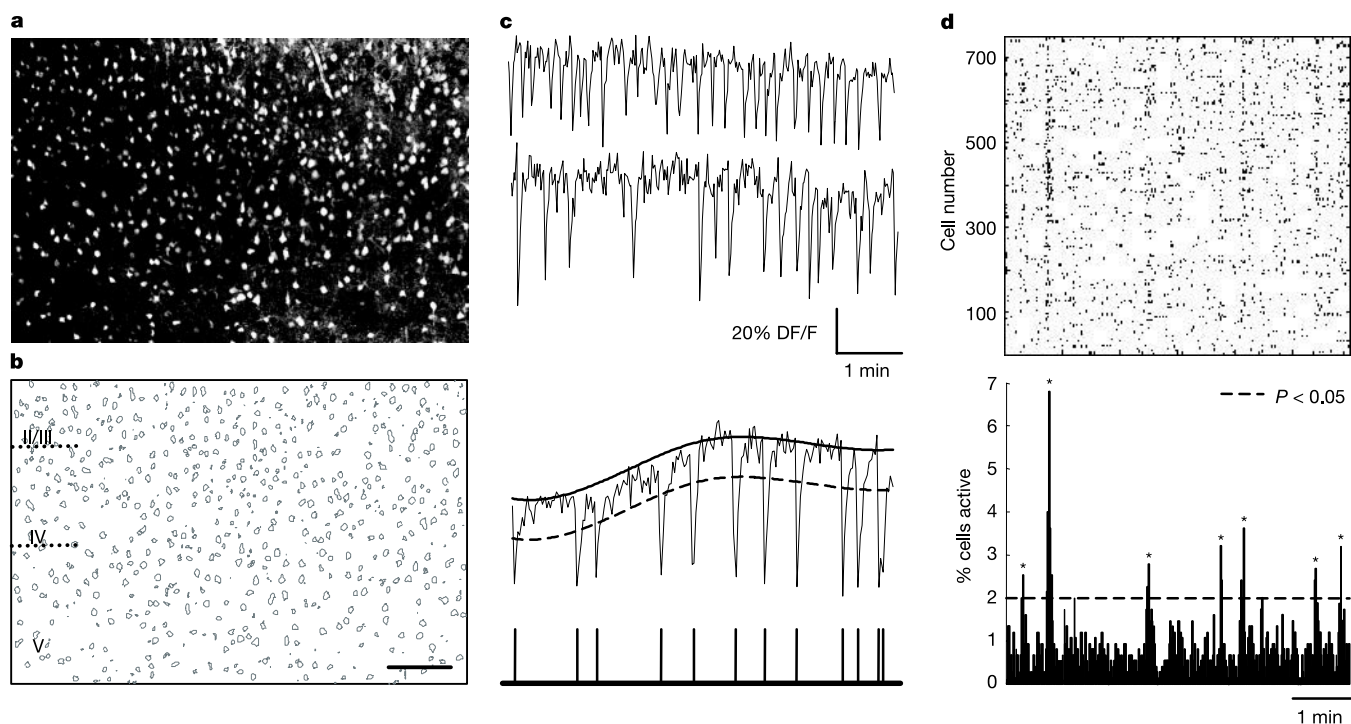


Figure 1 Two-photon calcium imaging of persistent activity in large neuronal populations. **a**, Two-photon calcium-fluorescence image taken at low magnification ($\times 20$) of a slice from area V1. The image is an average of 300 consecutive frames. Time resolution, 1.2 s per frame. For a sample movie, see Supplementary Information. **b**, Contours of the 754 cells from the fluorescence image were detected by an automatic procedure. Scale bar, 100 μ m. **c**, Top: representative traces of calcium transients. Time resolution,

1.2 s per frame. Bottom: example of calcium-transient detection. **d**, Top: representative rasterplot in which each row represents a single cell, and each mark a detected calcium transient. Bottom: histogram representing the percentage of cells active at each frame of the movie. Asterisks mark significant peaks of synchrony. Time resolution, 1.2 s per frame.

the patterns were found to be in one of the spatially stereotyped categories (clusters, layers or columns). This percentage increased gradually to 24% throughout the duration of the onset, and decreased gradually throughout the duration of the offset of the events ($n = 228$ events; not shown). Thus, the peaks of synchronous activity corresponded to maximal spatial organization of the coactive cells.

In most movies, more than one peak of synchronous activity was detected, and some cells were repeatedly active at several peaks (Fig. 3A). Specifically, 17% of the spatial patterns of coactive cells had a significant overlap ($P < 0.05$) with at least one other pattern at a peak of synchrony ($n = 228$ events). On average, $4.9 \pm 0.1\%$ of the coactive cells were repeatedly coactive at another peak. The overlap was more pronounced (30% of peaks, $9.5 \pm 0.2\%$ of cells) when cells active throughout the entire duration of the events were included. Thus, not only the neurons active at peaks of synchrony, but also the temporal sequences of coactive groups immediately before and after the peaks, tended to be significantly repetitive. We conclude that the spontaneous coactivations can repeatedly involve the same components of the network, following similar dynamical trajectories.

We investigated the physiological events responsible for the described patterns of activity by targeting cells active at peaks of synchrony online and performing current-clamp recordings (Fig. 4). Cells were also filled with biocytin and could thus be morphologically identified post hoc. Most cells active at peaks (71%, $n = 17$) were pyramidal neurons, whereas the rest ($n = 5$) were interneurons. Surprisingly, electrophysiological recordings revealed that the membrane potential of most coactive cells (96%, $n = 50$) fluctuated between two states (Fig. 4). This property was not observed in randomly patched cells ($n = 10$). UP and DOWN membrane potential states were -56 ± 1 mV and -65 ± 1 mV, respectively ($n = 33$). Durations of UP states ranged from about 60 ms to 30 s (Fig. 4B). Transitions to the UP state were fast and occurred within 58 ± 5 ms, whereas transitions to the DOWN state

could be fitted by a single exponential function with a time constant of 164 ± 46 ms. In 55% of cases, the UP state was suprathreshold, with a firing frequency of 21 ± 9 Hz and an average of 12 ± 4 spikes per state ($n = 12$). Suprathreshold UP states always produced detectable calcium transients (see Methods). In 44% of the neurons that showed suprathreshold UP states, low-frequency firing also occurred in the DOWN state (0.64 ± 0.18 Hz, $n = 8$ cells). However, such firing could not be detected in the imaging. In several cases, a cell participated a second time in a peak after it was patched, enabling us to perform simultaneous electrical and optical recordings during a synchronous event. During peaks of synchrony, cells did indeed exhibit a shift of the membrane potential to a depolarized UP state with sustained high-frequency firing (Fig. 4A). Thus, we conclude that cells involved in peaks of synchrony exhibit bimodal membrane potentials, and that synchrony occurs when several neurons simultaneously shift to suprathreshold UP states, defining a coactive neuronal ensemble.

Finally, we investigated the cellular and synaptic mechanisms underlying the occurrence of the peaks of synchrony. The intracellular incidence of UP states in cells participating in network UP states in control conditions was almost entirely prevented by the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)- and NMDA (*N*-methyl-D-aspartate)-receptor antagonists 2,3-dihydroxy-6-nitro-7-sulphamoylbenzo(f)quinoxaline (NBQX; $10 \mu\text{M}$) and MK801 ($10 \mu\text{M}$), respectively (UP state frequency decreased by 93%, $n = 19$ cells; Fig. 4B). Surprisingly, in a few electrophysiological recordings in the presence of blockers of glutamatergic transmission, UP states still occurred, even when all excitatory postsynaptic potentials (EPSPs) were blocked ($n = 10$ cells; Fig. 4B). This indicates that glutamatergic transmission might not be essential for the generation of UP states in some cells. In these cells, all UP state transitions were blocked when GABA (γ -aminobutyric acid)-mediated transmission was suppressed by the further addition of picrotoxin ($100 \mu\text{M}$, $n = 8$ cells), an antagonist of GABA_A receptors. Thus, we propose that UP states are synaptically

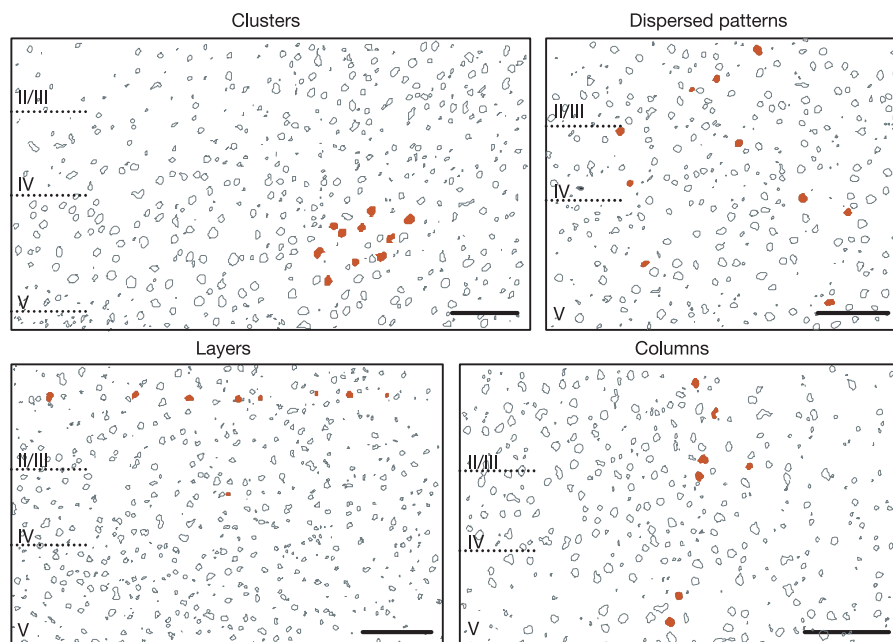


Figure 2 Spatial structure of synchronous neuronal ensembles. Representative contour plots of the four types of spatial arrangement of cells coactive at peaks of synchrony. Red contours indicate cells active at the movie frame corresponding to the peak of synchrony.

Patterns were identified and classified using Monte Carlo simulations (see Methods). The dispersed pattern illustrated was detected in the movie provided in Supplementary Information. Scale bar, $100 \mu\text{m}$.

driven by a combination of both glutamatergic and GABAergic transmission²⁴.

Two-photon calcium imaging of large neuronal populations, online reconstruction and analysis of individual signals, and intracellular recordings of targeted cells have enabled us to study the spatiotemporal dynamics of spontaneous cortical activity and to correlate it with intracellular responses. In the absence of any stimulation, networks in cortical slices exhibit characteristic patterns of synchronous activity. Coactivations occur during simultaneous transitions to membrane potential UP states. Coactive ensembles are often dispersed throughout different layers, but can be spatially organized in ways that reflect the underlying cortical architecture. The microcircuits of coactive cells participating in UP states are composed of small numbers of cells. Thus, these states differ from 'global' UP and DOWN states observed in other preparations^{11,13,16}, although the single-cell activity underlying them shares properties with the membrane fluctuations we report. Rather, the events we describe ('cortical flashes') are more reminiscent of the persistent activity states recorded in sparse populations

of cells during working memory tasks in awake monkeys^{17,18,25}.

Intriguingly, the spontaneous peaks of synchrony studied here follow recurrent stereotyped spatiotemporal dynamics (Fig. 3). This seems to be analogous to the dynamical behaviour of feedback neural networks, which converge into 'attractors', theoretically defined as stable states in network dynamics²². These emergent stable states could represent the solution to a specific computation²², and have been used to model orientation selectivity²⁶, eye position stability²¹ and working memory^{19,20}.

Specifically, network UP states persist on timescales much longer than those of the feedforward information flow through the cortex², indicating that they are dynamically stable (Fig. 4). Furthermore, they represent repeatable patterns of coincident firing, often following dynamical trajectories involving the same subcircuits (Fig. 3). Thus, network UP states are unexpectedly probable, which implies that they are preferred states of the network. Hence, the build-up of synchrony that we image would correspond to the descent into an attractor basin (UP state). Finally, we observe a multiplicity of discrete patterns, where different collections of neurons are involved

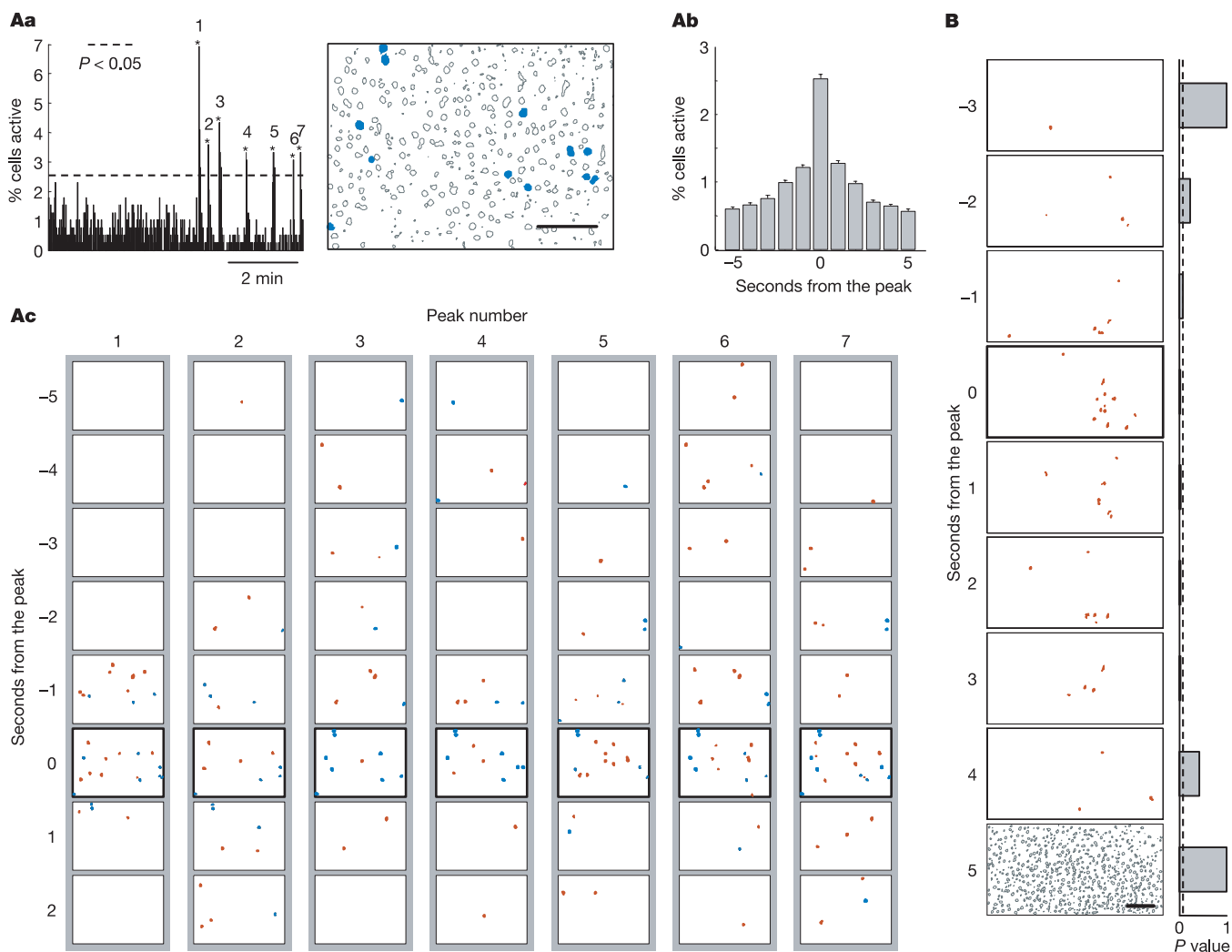


Figure 3 Spatiotemporal dynamics of synchronous activity. **Aa**, Repetitive activation of neuronal ensembles. Representative example showing multiple peaks of synchrony (asterisks). Blue contours indicate the cells repeatedly active at every peak. Time resolution, 1.1 s per frame. **Ab**, Average kinetics of peaks of synchrony from all movies ($n = 642$). **Ac**, Sequences of consecutive movie frames before and after each peak of

synchrony (bold) in **Aa**. Only active cells are indicated. The cells active at every peak in **Aa** are indicated in blue. **B**, Frames of another movie before and after a peak of synchrony (0, bold). Red contours indicate active cells. All imaged cells are shown at frame 5. The bar graph indicates the significance of spatial organization. Scale bar, 100 μm .

in different patterns, and we find that single cells can be active in more than one of them, confirming theoretical predictions for attractor dynamics²². An alternative interpretation of our results is that we are imaging the activity of 'reporter neurons'; that is, postsynaptic cells driven in a feedforward fashion by 'hidden' attractor states. This possibility, however, still requires attractors to be present in the network.

We conclude that 'cortical flashes' are dynamical attractors, and that the isolated cortical microcircuit can not only generate them, but also produce a rich variety with specific and discrete anatomical patterns. Our data are consistent with the hypothesis that a principal function of the highly recurrent neocortical networks is to generate persistent activity that might represent mental states^{3,7,27}. □

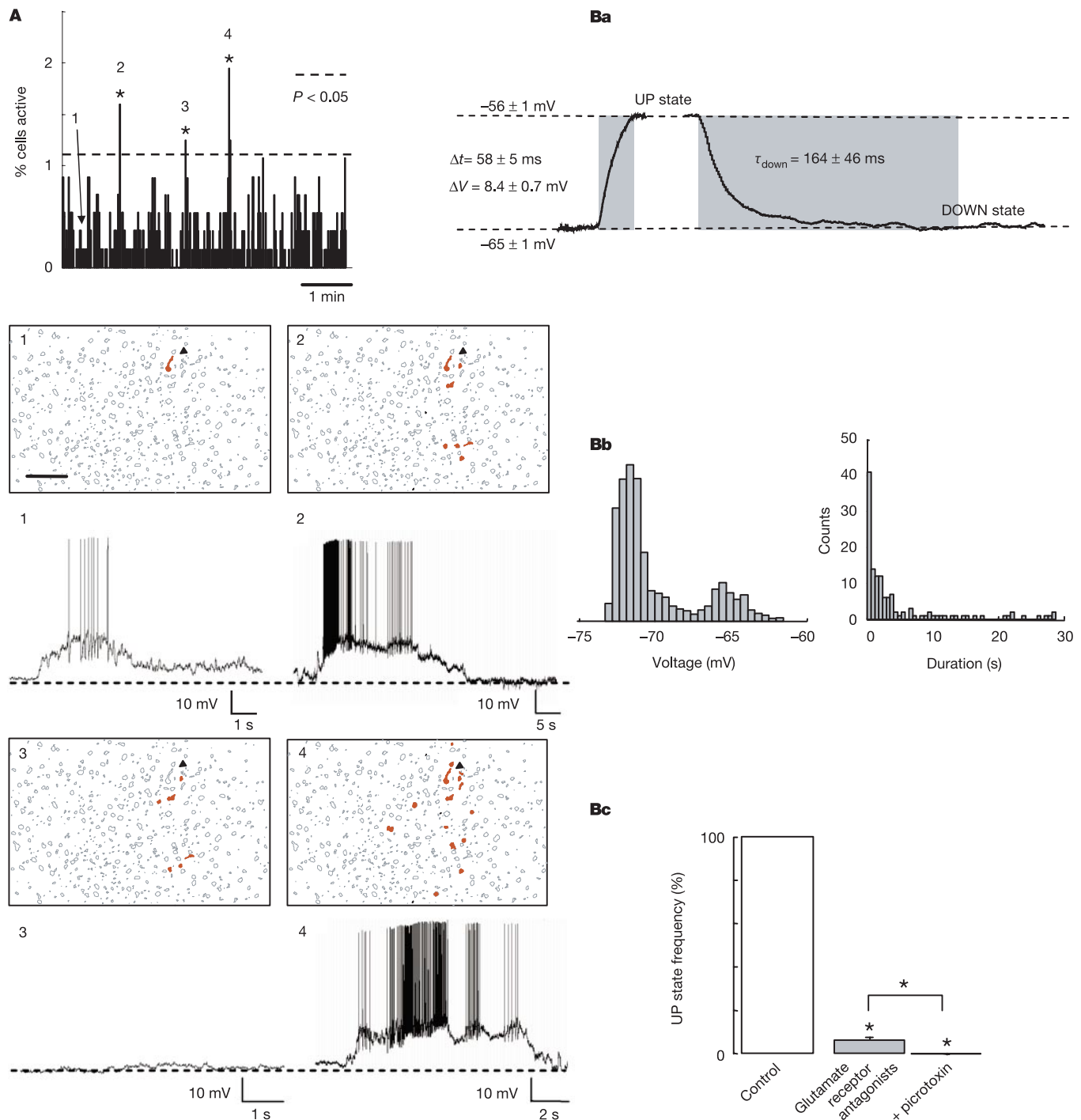


Figure 4 Synchronous activity is mediated by network UP states. **A**, Simultaneous imaging of network activity and current-clamp recording of a pyramidal cell (arrow) active at multiple peaks of synchrony (asterisks). Time resolution, 1.2 s per frame. Red contours indicate cells active at time points 1–4. Optical signals corresponded to shifts of the membrane potential to an UP state ($I = 0$ pA, $V_{\text{down}} = -70$ mV). Scale bar, 100 μ m.

Ba, Average kinetics of UP states ($n = 133$). Δt , length of onset; τ_{down} , time constant of average offset. **Bb**, Left: membrane potential distribution of a representative example. Right: distribution of UP state durations. **Bc**, Relative frequency of UP states in the presence of glutamatergic antagonists and after the further addition of the GABA receptor antagonist picrotoxin. Asterisks, $P < 0.05$.

Methods

Slice preparation and two-photon imaging

Slices (thickness, 300–350 μm) were prepared from visual cortices of P14–21 C57BL/6 mice using a Leica VT 1000S tissue slicer with ice-cold oxygenated modified aCSF that included 0.5 mM CaCl_2 and 7 mM MgSO_4 , in which NaCl was replaced by an equimolar concentration of sucrose. For AM-loading, slices were incubated in a small vial containing 2.5 ml oxygenated aCSF with 25 μl 1 mM Fura2-AM solution (Molecular Probes; in 100% dimethylsulphoxide) for 20–30 min. Slices were incubated in the dark, and incubation solution maintained at 35–37 °C. Experiments were performed at 35–37 °C with aCSF containing (in mM) 123 NaCl, 3 KCl, 26 NaHCO_3 , 1 NaH_2PO_4 , 2 CaCl_2 , 2 MgSO_4 and 10 dextrose, which was continuously aerated with 95% O_2 , 5% CO_2 . Imaging was carried out with a custom-made two-photon laser scanning microscope²⁸ consisting of a modified Fluoview (Olympus) confocal microscope with a Mira Ti:sapphire laser (Coherent) providing 130 fs pulses at 75 MHz, pumped by a solid-state source (Verdi, Coherent). Slices were imaged using a low-magnification, high-numerical-aperture objective ($\times 20$, NA-0.95, Olympus). Frames were acquired at 0.9–1.6 s per frame. Size of the imaged field typically $\sim 400 \times 700 \mu\text{m}$.

Electrophysiology

Cells were held in current clamp using a patch-clamp amplifier (BVC-700, Dagan Corporation). Pipette solution contained 130 mM potassium methylsulphate, 5 mM KCl, 5 mM NaCl, 10 mM HEPES, 2.5 mM Mg-ATP, 0.3 mM GTP, 30 μM Fura-2 pentapotassium salt (Molecular Probes) and 0.5% biocytin. The osmolarity was 265–275 mosM, pH 7.3. Microelectrodes had a resistance of 4–10 M Ω . Recordings were digitized online (10 kHz) with an A/D board (Instrutech) and analysed offline with MiniAnalysis 5.1 (Synaptosoft, Inc.). Neurons were filled with biocytin during recording and processed for morphological identification.

AMPA receptors were blocked by 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; 10 μM), 6,7-dinitro-2,3-quinoxalinedione (DNQX; 10 μM) or NBQX (10 μM). NMDA receptors were blocked by D-2-amino-5-phosphonopentanoate (AP5; 40 μM) or MK801 (10 μM). As we did not observe any significant differences in the effects of these antagonists, the data were pooled for further analysis. CNQX, DNQX, NBQX, picrotoxin and AP5 were purchased from Sigma; MK801 was obtained from Tocris.

Analysis

Online analysis was performed with custom-made software in ImageJ (National Institutes of Health) and Matlab (MathWorks). To detect calcium signals from imaged cells, we automatically identified loaded entities in images of the slices, and measured the average fluorescence of these entities as a function of time. To detect individual optical events, the baseline of each trace was calculated by smoothing the trace with a high-order Hanning filter (window length, 6–7 s). We estimated the amount of noise by subtracting the baseline from the trace and measuring the standard deviation of the values of all points above the baseline. We then divided the trace by this noise level to normalize it, and considered the threshold of 1 unit below the baseline. Five action potentials were usually necessary to produce a calcium transient of this magnitude. Event times were marked at the local minima of the trace that occurred below the threshold level (Fig. 1c).

To identify peaks of synchronous activity that included more cells than expected by chance, we used interval reshuffling (randomly reordering of intervals between events for each cell) to create sets of surrogate event sequences¹². Reshuffling was carried out 1,000 times for each movie, and a surrogate histogram was constructed for each reshuffling. The threshold corresponding to a significance level of $P < 0.05$ was estimated as the number of coactive cells exceeded in a single frame in only 5% of these histograms. We then examined each peak that was significant at $P < 0.05$, and manually rejected traces where noise was erroneously detected as an event. Peaks of synchronous activity that remained significant were selected for further analysis. For each peak of synchrony, we sought to determine whether the arrangement of cells involved could be classified into one of three spatial patterns (clusters, layers or columns) with confidence. We characterized each pattern of coactive cells by three measurements: the average distance between cell locations and their centroid, the average length of the projections of these distances in the direction parallel to the pia, and the average length of the projections in the direction perpendicular to the pia. For each pattern, we created 1,000 surrogate arrangements of cells using Monte Carlo simulations, choosing the same number of cells at random from the slice each time. All three measurements were made for each of the surrogate patterns, and the significance of the spatial arrangement was determined by directly comparing the resulting distributions with the original measurements. If distance projections were significantly small ($P < 0.05$) only in the direction parallel or perpendicular to the pia, the pattern was classified as a column or a layer, respectively. If the unprojected distances were significantly small, the pattern was classified as a cluster.

Received 16 December 2002; accepted 25 March 2003; doi:10.1038/nature01614.

- Mountcastle, V. B. *Perceptual Neuroscience: The Cerebral Cortex* (Harvard Univ. Press, Cambridge, Massachusetts, 1998).
- Hübner, D. H. & Wiesel, T. N. Functional architecture of the macaque monkey visual cortex. *Proc. R. Soc. Lond. B* **198**, 1–59 (1977).
- Lorente de Nó, R. Analysis of the activity of the chains of internuncial neurons. *J. Neurophysiol.* **1**, 207–244 (1938).
- Gilbert, C. & Wiesel, T. N. Morphology and intracortical projections of functionally characterised neurons in the cat visual cortex. *Nature* **280**, 120–125 (1979).
- Douglas, R. J. & Martin, K. A. C. In *The Synaptic Organization of the Brain* (ed. Shepherd, G. M.) 459–511 (Oxford Univ. Press, Oxford, 1998).
- Somogyi, P., Tamas, G., Lujan, R. & Buhl, E. Salient features of synaptic organisation in the cerebral

cortex. *Brain Res. Brain Res. Rev.* **26**, 113–135 (1998).

- Llinás, R. *I of the Vortex: From Neurons to Self* (MIT Press, Cambridge, Massachusetts, 2002).
- Creutzfeldt, O. *Cortex Cerebri* (Oxford Univ. Press, Oxford, 1995).
- Steriade, M., Contreras, D., Curro, D. R. & Nunez, A. The slow (<1 Hz) oscillation in reticular thalamic and thalamocortical neurons: scenario of sleep rhythm generation in interacting thalamic and neocortical networks. *J. Neurosci.* **13**, 3284–3299 (1993).
- Tsodyks, M., Kenet, T., Grinvald, A. & Arieli, A. Linking spontaneous activity of single cortical neurons and the underlying functional architecture. *Science* **286**, 1943–1946 (1999).
- Sanchez-Vives, M. & McCormick, D. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neurosci.* **3**, 1027–1034 (2000).
- Mao, B. Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C. & Yuste, R. Dynamics of spontaneous activity in neocortical slices. *Neuron* **32**, 883–898 (2001).
- Wilson, C. J. & Groves, P. M. Spontaneous firing patterns of identified spiny neurons in the rat neostriatum. *Brain Res.* **220**, 67–80 (1981).
- Cowan, R. L. & Wilson, C. J. Spontaneous firing patterns and axonal projections of single corticostriatal neurons in the rat medial agranular cortex. *J. Neurophysiol.* **71**, 17–32 (1994).
- Steriade, M., Nuñez, A. & Amzica, F. A novel slow (~1 Hz) oscillation of neocortical neurons *in vivo*: depolarizing and hyperpolarizing components. *J. Neurosci.* **13**, 3252–3265 (1993).
- Anderson, J., Lampl, I., Reichova, I., Carandini, M. & Ferster, D. Stimulus dependence of two-state fluctuations of membrane potential in cat visual cortex. *Nature Neurosci.* **3**, 617–621 (2000).
- Fuster, J. M. The prefrontal cortex and its relation to behavior. *Prog. Brain Res.* **87**, 201–211 (1991).
- Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* **14**, 477–485 (1995).
- Durstewitz, D., Seamans, J. K. & Sejnowski, T. J. Neurocomputational models of working memory. *Nature Neurosci.* **3** suppl., 1184–1191 (2000).
- Wang, X. J. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* **24**, 455–463 (2001).
- Seung, H. S., Lee, D. D., Reis, B. Y. & Tank, D. W. Stability of the memory of eye position in a recurrent network of conductance-based model neurons. *Neuron* **26**, 259–271 (2000).
- Hopfield, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl Acad. Sci. USA* **79**, 2554–2558 (1982).
- Badea, T., Goldberg, J., Mao, B. Q. & Yuste, R. Calcium imaging of epileptiform events with single-cell resolution. *J. Neurobiol.* **48**, 215–227 (2001).
- Shu, Y., Hasenstaub, A. & McCormick, D. A. Turning on and off recurrent balanced cortical activity. *Nature* **423**, 288–293 (2003).
- Constantinidis, C., Franowicz, M. & Goldman-Rakic, P. Coding specificity in cortical microcircuits: a multiple-electrode analysis of primate prefrontal cortex. *J. Neurosci.* **21**, 3646–3655 (2001).
- Ben-Yishai, R., Lev Bar-Or, R. & Sompolinsky, H. Orientation tuning by recurrent neural networks in visual cortex. *Proc. Natl Acad. Sci. USA* **92**, 3844–3848 (1995).
- Hebb, D. O. *The Organization of Behaviour* (Wiley, New York, 1949).
- Majewska, A., Yiu, G. & Yuste, R. A custom-made two-photon microscope and deconvolution system. *Pflugers Arch.* **441**, 398–409 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank D. McCormick for advice as well as J. Hopfield and members of our laboratory for comments. Supported by NEI, NINDS, FRM, the Human Frontier Science Program, the New York STAR Center for High Resolution Imaging of Functional Neural Circuits and the John Merck Fund.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.C. (rcosart@biology.columbia.edu).

Turning on and off recurrent balanced cortical activity

Yousheng Shu, Andrea Hasenstaub & David A. McCormick

Department of Neurobiology, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06510, USA

The vast majority of synaptic connections onto neurons in the cerebral cortex arise from other cortical neurons, both excitatory and inhibitory, forming local and distant ‘recurrent’ networks. Although this is a basic theme of cortical organization, its study has been limited largely to theoretical investigations, which predict that local recurrent networks show a proportionality or balance between recurrent excitation and inhibition, allowing the generation of stable periods of activity^{1–5}. This recurrent activity might underlie such diverse operations as short-term memory^{4,6,7}, the modulation of neuronal excitability with

attention^{8,9}, and the generation of spontaneous activity during sleep^{5,10–14}. Here we show that local cortical circuits do indeed operate through a proportional balance of excitation and inhibition generated through local recurrent connections, and that the operation of such circuits can generate self-sustaining activity that can be turned on and off by synaptic inputs. These results confirm the long-hypothesized role of recurrent activity as a basic operation of the cerebral cortex.

Maintaining slices of ferret prefrontal and occipital cortex *in vitro* resulted in the spontaneous and periodic (once every 3–7 s) generation of synaptic and action potential activity for periods of 1–3 s in cells recorded from layers II–VI (Fig. 1a); this activity was initiated earliest and most intensely in layer V¹³. Extracellular recordings from single layer-V neurons revealed that during the period of neuronal activity (the ‘UP’ state), these cells fired between 2 and 47 Hz (mean, 17.1 ± 11.1 Hz; $n = 32$). During the periods of relative quiescence (the ‘DOWN’ state), 18 of these cells did not discharge spontaneously, whereas the other 14 discharged at an average rate of 3.6 ± 3.0 Hz. These alternating UP and DOWN states, which are characterized by the participation of every neuron recorded, closely resemble those occurring *in vivo* during slow-wave sleep and anaesthesia^{10–14}. Intracellularly, each UP state *in vitro* was associated with a square or trapezoid-shaped 4–10-mV depolarization of pyramidal neurons and fast-spiking (FS) interneurons in synchrony with nearby multiple-unit activity (Fig. 1). The sustained activity of the UP state was shown to be generated through network mechanisms, as it was associated with large barrages of synaptic potentials, and because preventing the generation of action potentials by hyperpolarizing the neuron by 10–30 mV did not affect the duration of the UP state (normal, 1.7 ± 0.3 s; hyperpolarized, 1.7 ± 0.4 s; $n = 16$ cells) or its rate of recurrence (normal, 0.2 ± 0.1 Hz; hyperpolarized, 0.2 ± 0.1 Hz), as has also been observed *in vivo*^{10–14}. Bath application of the non-NMDA (*N*-methyl-D-aspartate) glutamate ionotropic receptor antagonist 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX; 5–10 μ M; $n = 8$) or the NMDA receptor antagonist DL-2-amino-5-phosphonopentanoate (DL-AP5; 10–25 μ M; $n = 7$) completely blocked the generation of recurrent activity in extracellular multiple-unit recordings, although a minority of neurons continued to exhibit low-level spontaneous action potential activity (not shown; see ref. 5). Block of GABA_A (γ -aminobutyric acid type A) receptors with bath-applied picrotoxin (50 μ M; $n = 8$) transformed slow oscillations into seizure-like activity¹³ (see Supplementary Information; for more information, see <http://www.mccormicklab.org>).

Depolarizing current pulses did not result in the generation of ‘after discharges’ in single cortical neurons ($n = 25$), as has been observed following the artificial stimulation of muscarinic receptors¹⁵. Moreover, the interspike intervals of regular-spiking neurons during our UP states were highly variable (average coefficient of variation, 1.74; $n = 6$; for example, see Fig. 4b). Highly variable spike timing is typical of cortical activity *in vivo* under a variety of conditions³; for example, in prefrontal cortical neurons *in vivo* during working memory tasks^{16,17}. The action potential activity during the UP state is markedly different from the very regular discharge generated intrinsically by afterdepolarizing currents¹⁵.

We next determined whether or not UP states can also be initiated and terminated through the activation of synaptic pathways. Activation of neuronal elements with single-shock electrical stimulation applied either locally (within layer V), within layers II/III, or in the white matter below the recording site, could initiate the UP state (Fig. 1b), with the duration of the UP state depending on the intensity of the stimulus. Increasing the intensity of the shock decreased the duration of the UP state, such that a strong electrical stimulus (greater than four times the threshold) resulted in only a brief (20–30 ms) burst of activity in the multiple-unit recording ($n = 6$; not shown). Interestingly, delivery of the same shock used to initiate the UP state could also terminate the recurrent activity,

depending on the time since the onset of activity and the strength of the stimulus ($n = 5$). Although relatively weak stimuli were unable to terminate the UP state (Fig. 1c, 60 μ A), moderately strong stimuli could stop it, but only after a delay, the duration of which depended on the intensity of the stimulus (Fig. 1c, 180 μ A). Strong stimuli could terminate the UP state at intervals as short as 100 ms, even in a slice whose UP state naturally lasted for 1–3 s (Fig. 1c, 360 μ A).

We next addressed two questions: how do local cortical networks generate this recurrent activity, and how do afferent inputs start and stop it? Examination of the average postsynaptic currents (PSCs) arriving in layer-V pyramidal neurons at different membrane potentials with single-electrode voltage clamp revealed that the UP state was associated with PSC barrages (Fig. 2a) whose average reversal potential¹⁸ was -30.1 ± 9.8 mV ($n = 6$; Fig. 2b). The reversal potential varied remarkably little over most of the UP period (average standard deviation during the UP state, 2.4 ± 1.6 mV). In most cells, the onset of the UP period was associated with a more depolarized reversal potential, which rapidly moved towards, and sometimes undershot, the average reversal potential of the rest of the UP period (Fig. 2b).

Examination of the average input conductance of the cell during the generation of the UP state revealed a steady increase in conductance followed by a steady decrease (Fig. 2c). Calculating the theoretical contributions of excitatory (G_e) and inhibitory (G_i) conductances revealed that these increase and decrease in close association with each other (Fig. 2c; see Methods)¹⁸. Plotting the inhibitory conductance as a function of excitatory conductance revealed a nearly linear relationship (Fig. 2d), with an average slope of 0.68 for G_i/G_e ($n = 5$), indicating that the amplitude of recurrent inhibition within the local network was precisely related to the amplitude of recurrent excitation. Similarly, both the inhibitory and excitatory conductances were strongly correlated with the amplitude of multiple-unit activity during the UP state (Fig. 2e; see Methods). These results indicate that the local generation of

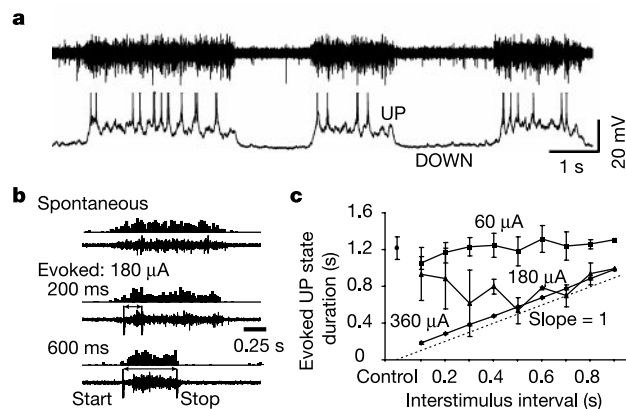


Figure 1 Prefrontal cortical slices generate periods of recurrent activity both spontaneously and in response to electrical stimulation of the neuropil. **a**, Simultaneous intracellular (pyramidal cell; bottom) and extracellular (top) recording in layer V of the prefrontal cortex during the spontaneous generation of recurrent activity. Transitions between states were rapid, typically complete within <150 ms. **b**, Extracellular multiple-unit recording showing that the stimulation of afferent pathways and nearby cells can both start and stop the generation of recurrent activity. A multiple-unit recording and a spike histogram are shown. The first electrical stimulus to the white matter is referred to as the start stimulus, and the second is referred to as the stop. **c**, Delivery of a 60- μ A stimulus does not consistently affect the duration of the UP state, although it started it in 69% of the trials. A 180- μ A stimulus does stop the UP state, but only if it is delivered after a few hundred milliseconds from the onset of the sustained activity (and elicits the UP state on 92% of the trials). Delivery of a strong stimulus (360 μ A) stops the UP state at all intervals (and always starts the UP state). Movies of the recurrent activity can be viewed in Supplementary Information.

sustained activity in the cerebral cortex is associated with a striking proportionality (balance) of recurrent excitatory and inhibitory conductances, even when the total conductance is undergoing large changes. Since the membrane potential during the recurrent activity state is near -60 to -55 mV, average excitatory synaptic currents

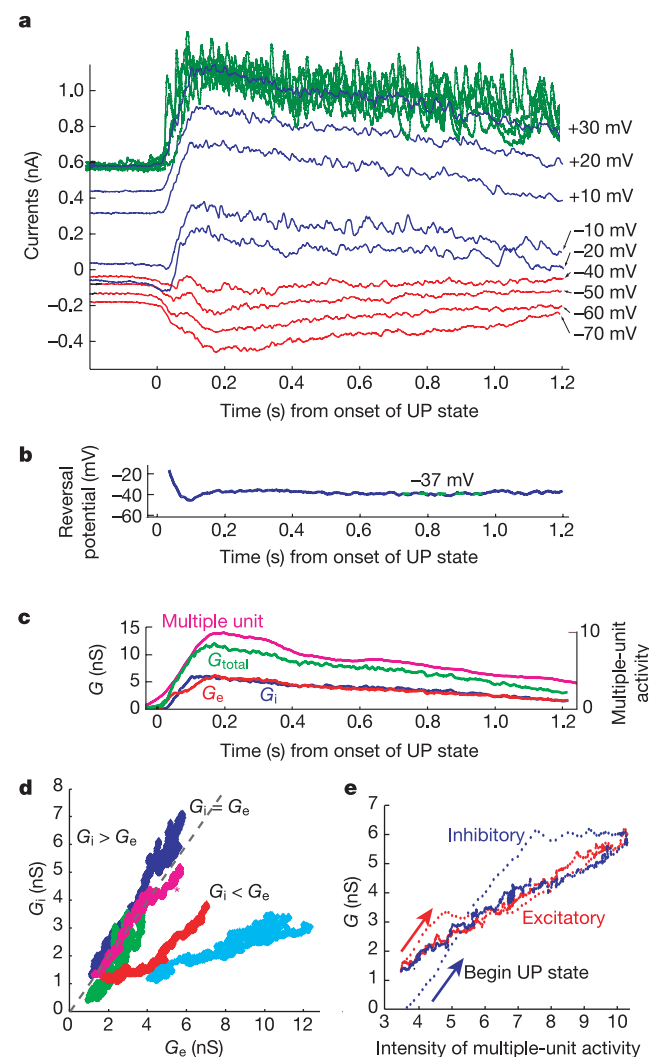


Figure 2 Recurrent activity is generated by a balanced barrage of IPSPs and EPSPs. **a**, Average currents during the UP state under voltage clamp. Each trace is an average of 9–17 trials. Several raw traces at +30 mV are shown with the average for comparison. The average synaptic currents reverse at around -37 mV in this cell. The electrode contained 2 M caesium acetate and 50 mM QX-314 to minimize the contribution of K^+ and Na^+ currents. Similar results were obtained with electrodes containing potassium acetate only (see Supplementary Information). **b**, Calculation of the reversal potential of the average synaptic currents over time. **c**, Additional conductance during the UP state. Illustration of the amplitude time course during the UP state of the average multiple-unit activity, the total increase in conductance (as measured by the change in slope of the I - V plot), and the calculated conductance of excitatory and inhibitory currents²². **d**, Plot of the inhibitory conductance as a function of the excitatory conductance in five different neurons. **e**, Synaptic conductance tracks network activity. The relationship is shown between the average intensity of neuronal activity in the local network, as measured by multiple-unit activity (see Methods), and the amplitude of the calculated excitatory and inhibitory conductances. In this cell, as the network transitioned into the UP state, the inhibitory conductance lagged behind the excitatory conductance. After the onset of the UP state, the excitatory and inhibitory conductances were proportional and strongly correlated with the intensity of activity in the nearby ($<100 \mu\text{m}$) layer-V multiple-unit recording.

will be considerably larger than inhibitory ones. Other currents that contribute to the generation of a stable membrane potential include outward leak currents, as well as K^+ currents activated during the depolarization of the UP state.

To address the possibility that this balance may be disrupted by local stimuli that are successful in stopping the recurrent activity, we examined the activity of pyramidal and FS GABAergic neurons in our two stimulation protocol (Figs 3 and 4). Stimulus-evoked UP states were characterized by a prolonged, synaptically mediated depolarization in both pyramidal and FS GABAergic neurons. A second stimulus, identical to the first, resulted in excitation of both pyramidal cells and FS interneurons, followed by the withdrawal of postsynaptic potential (PSP) barrages. The starting stimulus resulted in 2.1 ± 0.48 ($n = 6$) spikes in FS interneurons, whereas the stopping stimulus activated 3.6 ± 0.69 spikes ($P < 0.01$; Fig. 3a). The instantaneous discharge frequency of FS interneurons was significantly higher for the stop than for the start stimulus ($n = 5$; Fig. 3a). These effects appear to result largely from the depolarization associated with the UP state, because examination of the PSP barrages evoked by the start and stop stimuli, after hyperpolarization to prevent the generation of action potentials, failed to reveal marked differences in the amplitude time course of these events ($n = 3$; Fig. 3b). In addition, depolarization of FS interneurons by an amount similar to that of the UP state resulted in a similar increase in action potential response to the start stimulus ($n = 3$; not shown). This result is consistent with the proposal that activation of inhibitory interneurons might cause the cessation of recurrent activity in cortical networks⁷.

Examination of action potential responses in pyramidal cells also revealed a significant increase in responsiveness during the UP state (Fig. 4a, b). The starting stimulus resulted in an average of 0.45 ± 0.32 spikes in regular-spiking pyramidal cells at a latency of 7.02 ± 2.48 ms ($n = 6$), whereas the stopping stimulus resulted in an average of 0.80 ± 0.09 spikes per stimulus ($P < 0.01$) at a shorter latency of 5.43 ± 0.94 ms ($P < 0.05$; Fig. 4a, b). The increased responsiveness of both FS interneurons and pyramidal cells during the UP state was corroborated by examining the peak amplitude of the PSCs evoked while holding a pyramidal cell near either -75 or 0 mV, to isolate excitatory PSCs (EPSCs) and inhibitory PSCs (IPSCs), respectively. At 0 mV, the stopping stimu-

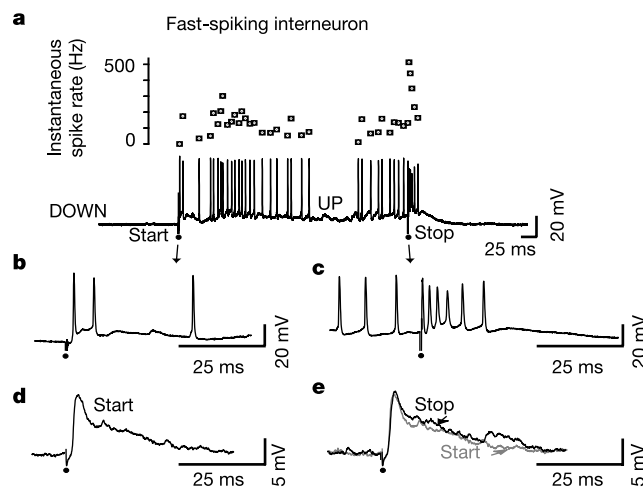


Figure 3 The UP state enhances neuronal responses to PSPs in FS interneurons. **a**, Intracellular recording from an FS interneuron and its response to two single-shock stimuli delivered to the white matter. **b**, **c**, Expansions of the FS interneuron response to the start and stop stimuli. **d**, **e**, Hyperpolarization of the cell so that action potentials were not generated revealed the underlying PSP barrages, which are similar for start and stop stimuli (see overlaid traces in **e**).

lus evoked an IPSC that was significantly larger than that evoked by the starting stimulus (peak amplitude for stop stimulus, 1.03 ± 0.16 nA; peak for start stimulus, 0.83 ± 0.15 nA; $P < 0.01$; $n = 13$). Similarly, the peak amplitude of the stop-stimulus-evoked EPSC at -75 to -80 mV was larger than that of the start-stimulus-evoked EPSC (stop stimulus, 0.85 ± 0.51 nA; start stimulus, 0.64 ± 0.48 nA; $P < 0.01$; $n = 6$). See Fig. 5b for an example of the increase in calculated excitatory and inhibitory conductance to the stop stimulus compared the start stimulus. The increase in the amplitude of evoked PSCs and in the number of spikes per stimulus during the UP state is probably caused by depolarization of the membrane potential of the recorded cell as well as the other cells in the network, although increased membrane potential fluctuations might also contribute⁹ (Y.S., A.H. and D.A.M., unpublished observations).

There is an important difference between the responses of pyramidal cells and FS interneurons to synaptic inputs. Both the start and stop stimuli evoked at most only single action potentials in pyramidal cells (Fig. 4a, b), whereas FS interneurons, especially during the UP state, responded to local stimulation with a prolonged discharge of spikes (Fig. 3a, b). This more prolonged increase in discharge rate in FS interneurons during the UP state might cause a negative shift in the reversal potential of the response to the stopping (compared with the starting) stimulus and, indeed, this was observed in a significant number of cells (binomial statistics, $P < 0.05$) 14–53 ms after activation of the synaptic inputs (Fig. 5a, c), and the average peak shift in the reversal potential was -10 ± 6.4 mV ($n = 10$).

The important role of activation of inhibition in the termination of recurrent activity is also suggested by the observation that increasing the amplitude of the electrical stimulus gradually increased the amplitude of the evoked period of inhibition after the initial excitation (for example, see the pause after the start stimulus in Fig. 4b)¹⁹. Increasing the start stimulus to the point at which no UP state was generated was associated with the activation of excitatory PSPs (EPSPs) followed by prominent inhibitory PSPs (IPSPs) in pyramidal neurons ($n = 6$; not shown)¹⁹.

Our results show that recurrent activity within the cerebral cortex can be generated though the activation of local excitatory and inhibitory networks, and that the cortex is well tuned to balance the amplitude of recurrent inhibition with that of excitation (as measured at the soma), resulting effectively in a 'reversal-potential

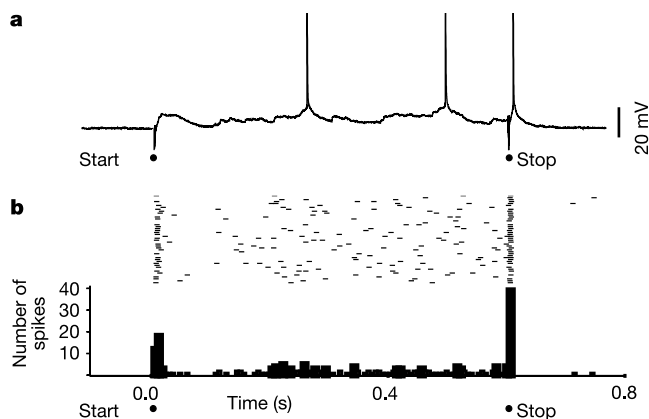


Figure 4 The UP state enhances neuronal responses to PSPs in pyramidal cells. **a**, Intracellular recording in a layer-V pyramidal cell illustrating that the UP state enhances the response to the activation of synaptic inputs. **b**, Rasters and peristimulus histogram showing that the UP state is associated with an increased response to the activation of synaptic inputs to this pyramidal cell ($n = 46$ trials).

clamp', even during large changes in total synaptic conductance. This basic feature of local cortical circuits arises directly from their architecture: pyramidal cells are interconnected in a positive-feedback manner with a negative-feedback loop operated by local GABAergic interneurons. Although the precise pattern of excitation and inhibition in cortical neurons depends crucially on the spatio-temporal patterns of afferent activity (and therefore may be dominated by inhibition, excitation or a balance of both^{18–21}), our results show that the operation of local circuits can generate activity that exhibits a proportional increase in feedback excitation and inhibition, keeping the network relatively in balance. Balancing the membrane potential of cortical cells near firing threshold has been proposed to be important in the generation of highly variable interspike intervals^{2,3}, persistent activity *in vivo*^{1,4,5,7}, and as a mechanism by which cortical networks can modulate (by changing the properties of background synaptic barrages) neuronal responsiveness according to 'context' (for example, attention, far-field receptive-field stimulation, and so on)^{9,22}.

Activation of synaptic inputs can either initiate or terminate this recurrent activity, depending in part on the history of the network. In the inactive state, moderate activation of synaptic inputs depolarizes both inhibitory and excitatory neurons to a degree that allows recurrent excitation to flip the network into the active state⁵. However, during the sustained discharge state, activation of synaptic inputs results in a period of inhibition more prolonged than that of excitation, presumably reflecting the ability of FS interneurons to generate longer periods of activity than pyramidal cells, a pattern that might result from the preponderance of excitatory input to FS neurons. The activation of inhibitory mechanisms, both synaptic and intrinsic, might 'tip the balance' towards the reduced-activity state, and therefore result in the deactivation of recurrent activity⁷. This has been proposed to be a mechanism for the cessation of persistent activity *in vivo* on the successful completion of a working memory task^{7,17}. The increased probability of activation of an action potential in pyramidal cells by synaptic inputs during the generation of recurrent activity suggests that a synchronized refractory period might also contribute strongly to its

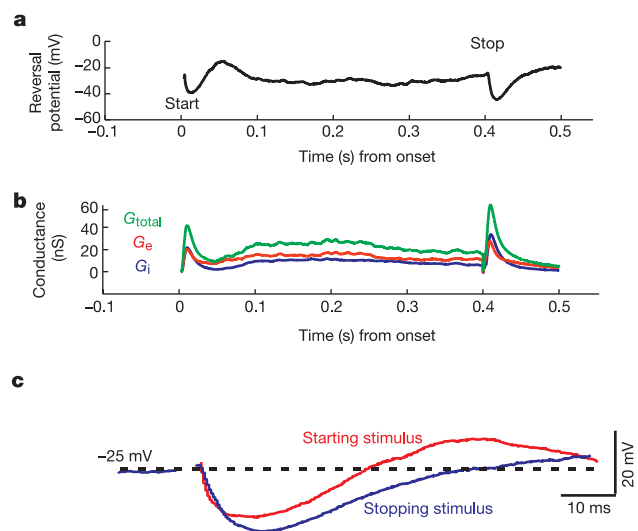


Figure 5 Reversal potential of evoked responses for the start and stop stimuli. **a**, Calculated reversal potential of the responses to start and stop stimuli as well as during the UP state. **b**, Calculated excitatory and inhibitory conductances that could yield the reversal potential of part **a**. The total conductance ($G_e + G_i$) is also shown. Note that the second stimulus causes a greater increase in conductance than the first. **c**, Overlap of the reversal potentials calculated for the start and stop stimuli. The response to the stop stimulus has a more hyperpolarized reversal potential.

cessation²³, as fewer pyramidal cells will be available to continue the activity following the stimulus.

We propose that a basic feature of cortical networks is the ability to generate multiple states of activity within local and long-distance recurrent networks, which in its simplest form is similar to a network 'flip-flop' involving nearly all neurons, that can be turned on or off by synaptic inputs, as we show here (Fig. 1). During sleep and in our slices, this is expressed as recurring oscillations of activity at about 0.1–0.5 Hz, and it might be important for the tuning of network processes¹². In the waking state, the cerebral cortex might generate self-sustained spontaneous 'background' activity, perhaps through mechanisms akin to a persistent UP state¹². Another example of sustained cortical activity is that which occurs during working memory tasks, in which neighbouring neurons often generate unrelated patterns of activity^{6,17}; this contrasts with the recurrent activity in our, but not in other, *in vitro* studies^{24,25}. In either case (sparse or dense neuronal participation), local cortical networks have a powerful mechanism at their disposal to quickly control the excitability and responsiveness of other cortical cells^{9,26}. The ability to generate self-sustaining periods of activity provides a computationally powerful mechanism by which cortical networks may solve a large variety of tasks^{27,28}. □

Methods

In vitro experiments were performed on 0.4-mm-thick slices, mainly from 2–4-month-old ferret prefrontal cortex (anterior to the Sylvian fissure), but also from occipital cortex (the periodic and recurrent activity studied here did not appear in slices from animals younger than approximately 2 months). The slices were maintained in an interface-style recording chamber at 35–36 °C (ref. 13). Slices were formed on a DSK Microslicer (Ted Pella Inc.) in a slice solution in which sucrose was used as a substitute for NaCl. Following transfer to the recording chamber, the slices were incubated in a 'traditional' slice solution containing (in mM): NaCl, 126; KCl, 2.5; MgSO₄, 2; NaHPO₄, 1.25; CaCl₂, 2; NaHCO₃, 26; dextrose, 10. This solution was aerated with 95% O₂, 5% CO₂ to a final pH of 7.4. After approximately 1 h, the slice solution was modified to be closer to that of normal cerebrospinal fluid, and contained 1 mM MgSO₄, 1 or 1.2 mM CaCl₂, and 3.5 mM KCl. Extracellular multiple-unit recordings were obtained with single electrodes (Frederick Haer Corporation). Simultaneous extracellular multiple-unit and intracellular recordings were performed in layer V after 2 h of recovery, with the electrodes within approximately 100 µm of each another. Intracellular recordings were carried out with bevelled sharp microelectrodes containing 2 M potassium acetate with resistances of 60–90 MΩ. Electrodes used for single-electrode voltage clamp contained 2 M caesium acetate to reduce K⁺ currents and 50 mM QX-314, which blocks voltage-dependent Na⁺ currents, the h-current, and GABA_B-receptor-mediated IPSPs. A few traces contained obvious intrinsically generated regenerative events, such as putative Ca²⁺ spikes at membrane potentials of around –40 mV, and were excluded from our analysis.

Extracellular and intracellular data were collected using Spike-2 software (Cambridge Electronic Design). To detect the onset of the UP state, the multiple-unit recording was rectified and smoothed to yield an outline of the increase and decrease in activity associated with the onset and offset of the UP state. Two thresholds were set, and the multiple-unit activity was required to rise above the highest threshold to be counted as an UP state. The onset and offset of the UP state were then determined by the crossings of the lower threshold, which was set at approximately 2 × baseline. Only UP states that were at least 0.5 s in duration were considered. This rectified and smoothed version of the multiple-unit activity was also used to form correlations between the average intensity of activity in the network and the calculated excitatory and inhibitory conductances in single neurons (Fig. 2e).

The increase in membrane conductance during the UP state compared with the DOWN state was measured using the method of Borg-Graham *et al.*¹⁸. PSCs occurring during the UP state were measured at several different membrane potentials under single electrode voltage clamp. All UP states in this analysis were longer than 1.2 s. Current–voltage (*I*–*V*) plots were formed every 0.5 ms for the DOWN and UP states, and the change in slope of *I*–*V* plots was taken as the change in membrane conductance during the UP state. The point of actual crossing of the *I*–*V* plot was taken as the reversal potential. Owing to the blocking action of QX-314, and because of the small conductance of GABA_B-receptor-mediated IPSPs, we did not include this conductance in our calculation of *G*_e and *G*_i¹⁸. We assume that our recording electrodes were located in or near the soma and, therefore, that the results we obtain are most relevant to synaptic integration at this compartment. Significant differences in the balance of excitation and inhibition in other regions of the neuron are likely to occur. During single-electrode voltage clamp, with a switching frequency of around 3,000 Hz, the headstage output of the Axoclamp-2B amplifier (Axon Instruments) was continuously monitored to ensure an adequate settling time. In cortical neurons, the reversal potential of EPSCs is around 0 mV, whereas for GABA_A-receptor-mediated IPSCs, it is around –75 mV (refs 29, 30). To confirm these numbers, as well as to test the limitations of our voltage-clamp protocol, we measured the reversal potential of electrical-stimulus-evoked EPSC barrages (following the block of GABA_A receptor inhibition with the bath (50 µM) or local (500 µM in micropipette) application of picrotoxin), as well as the reversal potential of evoked IPSCs after the block

of EPSCs with bath-applied CNQX (30 µM) and AP5 (50 µM). Following the block of IPSCs, evoked EPSC barrages were found to reverse at an average membrane potential of –6.2 ± 2.7 mV (*n* = 6) throughout the EPSC barrage, even though there were large changes in membrane conductance (see Supplementary Information). After the pharmacological block of EPSCs, the evoked IPSCs reversed at an average of –79 ± 1.4 mV (*n* = 6) throughout their time course (see Supplementary Information). In addition, the reversal potential of responses to the local application of glutamate (1 mM in the puffer pipette; 50 µM DL-AP5 and 100 µM picrotoxin in the bath to isolate AMPA (α-amino-3-hydroxy-5-methyl-4-isoxazole propionic acid)-receptor-mediated responses) to the apical dendrites of layer-V cells was near 0 mV, as expected (0.8 ± 1.5 mV; *n* = 6; see Supplementary Information). These results suggest that our voltage-clamp and reversal-potential measurement technique is reasonably accurate. If anything, our measures of the relative contribution of excitation and inhibition in single neurons may underestimate excitatory, relative to inhibitory, inputs, owing to the distribution of these different types of synapse. It should be noted that we use the term 'balance' to mean that excitatory and inhibitory inputs covary in a proportional manner.

Neuronal elements were activated locally with the delivery of single-shock electrical stimuli (duration, 100 µs; amplitude, 10–500 µA) through a concentric, bipolar electrode (Fredrick Haer Corporation) typically placed in the white matter near the border with layer VI, in line with the intracellularly recorded layer-V cell. Although this type of stimulus delivers an unphysiological level of synchrony among directly activated elements, it was a useful tool to investigate the possible mechanisms of initiation and termination of recurrent activity. Instances in which the recorded neuron exhibited antidromic activation were rare and were not included in the analysis. Periods that included stimulus artifacts were not analysed.

Received 12 December 2002; accepted 28 March 2003; doi:10.1038/nature01616.

- Wang, X. J. Synaptic reverberation underlying mnemonic persistent activity. *Trends Neurosci.* **24**, 455–463 (2001).
- van Vreeswijk, C. & Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural Comput.* **10**, 1321–1371 (1998).
- Shadlen, M. N. & Newsome, W. T. The variable discharge of cortical neurons: implications for connectivity, computation, and information coding. *J. Neurosci.* **18**, 3870–3896 (1998).
- Durstewitz, D., Seamans, J. K. & Sejnowski, T. J. Neurocomputational models of working memory. *Nature Neurosci.* **3**, 1184–1191 (2000).
- Compte, A., Sanchez-Vives, M. V., McCormick, D. A. & Wang, X. J. Cellular and network mechanisms of slow oscillatory activity (<1 Hz) in a cortical network model. *J. Neurophysiol.* **89**, 2707–2725 (2003).
- Fuster, J. M. *Memory in the Cerebral Cortex* (MIT press, Cambridge, Massachusetts, 1995).
- Brunel, N. & Wang, X. J. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J. Comput. Neurosci.* **11**, 63–85 (2001).
- Hahnloser, R. H. R., Douglas, R. J. & Hepp, K. Attentional recruitment of inter-areal recurrent networks for selective gain control. *Neural Comput.* **14**, 1669–1689 (2002).
- Chance, F. S., Abbott, L. F. & Reyes, A. D. Gain modulation from background synaptic input. *Neuron* **35**, 773–782 (2002).
- Steriade, M., Nunez, A. & Amzica, F. A novel slow (<1 Hz) oscillation of neocortical neurons *in vivo*: depolarizing and hyperpolarizing components. *J. Neurosci.* **13**, 3252–3265 (1993).
- Contreras, D., Timofeev, I. & Steriade, M. Mechanisms of long lasting hyperpolarizations underlying slow sleep oscillations in cat corticothalamic networks. *J. Physiol. (Lond.)* **494**, 251–264 (1996).
- Steriade, M., Timofeev, I. & Grenier, F. Natural waking and sleep states, a view from inside neocortical neurons. *J. Neurophysiol.* **85**, 1969–1985 (2001).
- Sanchez-Vives, M. V. & McCormick, D. A. Cellular and network mechanisms of rhythmic recurrent activity in neocortex. *Nature Neurosci.* **3**, 1027–1034 (2000).
- Cowan, R. L. & Wilson, C. J. Spontaneous firing patterns and axonal projections of single corticostriatal neurons in rat medial agranular cortex. *J. Neurophysiol.* **71**, 17–32 (1994).
- Egorov, A. V., Hamam, B. N., Franssen, E., Hasselmo, M. E. & Alonso, A. A. Graded persistent activity in entorhinal cortex neurons. *Nature* **420**, 173–178 (2002).
- Compte, A. *et al.* Temporal fluctuations of mnemonic persistent activity in prefrontal neurons of monkeys during a delayed response task. *J. Neurophysiol.* (submitted).
- Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
- Borg-Graham, L. J., Monier, C. & Frégnac, Y. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* **393**, 369–372 (1998).
- Hirsch, J. A. & Gilbert, C. D. Synaptic physiology of horizontal connections in cat's visual cortex. *J. Neurosci.* **11**, 1800–1809 (1991).
- Anderson, J. S., Carandini, M. & Ferster, D. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.* **84**, 909–926 (2000).
- Hirsch, J. A., Alonso, J. M., Reid, R. C. & Martinez, L. M. Synaptic integration in striate cortical simple cells. *J. Neurosci.* **18**, 9517–9528.
- Carandini, M. & Heeger, D. J. Summation and division by neurons in primary visual cortex. *Science* **264**, 1333–1336 (1994).
- Gutkin, B. S., Laing, C. R., Colby, C. L., Show, C. C. & Ermentrout, G. B. Turning on and off with excitation: the role of spike-timing asynchrony and synchrony in sustained neural activity. *J. Comput. Neurosci.* **11**, 121–134 (2001).
- Mao, B. Q., Hamzei-Sichani, F., Aronov, D., Froemke, R. C. & Yuste, R. Dynamics of spontaneous activity in neocortical slices. *Neuron* **32**, 883–898 (2001).
- Cossart, R., Aronov, D. & Yuste, R. Attractor dynamics of network UP states in the neocortex. *Nature* **423**, 283–288 (2003).
- Bernander, Ö., Douglas, R. J., Martin, K. A. C. & Koch, C. Synaptic background activity influences spatiotemporal integration in single pyramidal cells. *Proc. Natl Acad. Sci. USA* **88**, 11569–11573 (1991).
- Anderson, J. S., Lampl, I., Gillespie, D. C. & Ferster, D. The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science* **290**, 1968–1972 (2000).

28. Hahnloser, R. H., Sarpeshkar, R., Mahowald, M. A., Douglas, R. J. & Seung, H. S. Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* **405**, 947–951 (2000).
29. Stern, P., Edwards, F. A. & Sakmann, B. Fast and slow components of unitary EPSCs on stellate cells elicited by focal stimulation in slices of rat visual cortex. *J. Physiol. (Lond.)* **449**, 247–278 (1992).
30. Connors, B. W., Malenka, R. C. & Silva, L. R. Two inhibitory postsynaptic potentials, and GABA_A and GABA_B receptor-mediated responses in neocortex of rat and cat. *J. Physiol. (Lond.)* **406**, 443–468 (1988).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank R. Yuste for his comments. This work was supported by the National Institutes of Health (D.A.M.), the Human Frontier Science Program (D.A.M.), and by a fellowship from the Howard Hughes Institute (A.H.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.A.M. (david.mccormick@yale.edu).

Recurrent *de novo* point mutations in lamin A cause Hutchinson–Gilford progeria syndrome

Maria Eriksson*, W. Ted Brown†, Leslie B. Gordon‡, Michael W. Glynn§, Joel Singer||, Laura Scott||, Michael R. Erdos*, Christiane M. Robbins*, Tracy Y. Moses*, Peter Berglund¶, Amalia Dutra*, Evgenia Pak*, Sandra Durkin§, Antonei B. Csoka#, Michael Boehnke||, Thomas W. Glover§ & Francis S. Collins*

* National Human Genome Research Institute, and † Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

‡ Department of Human Genetics, New York State Institute for Basic Research in Developmental Disabilities, Staten Island, New York 10314, USA

§ Department of Anatomy and Cellular Biology, Tufts University School of Medicine, Boston, Massachusetts 02111, and Department of Pediatrics, Rhode Island Hospital, Providence, Rhode Island 02903, USA

|| Department of Human Genetics, and || Department of Biostatistics, University of Michigan, Ann Arbor, Michigan 48109, USA

Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, Providence, Rhode Island 02912, USA

Hutchinson–Gilford progeria syndrome (HGPS) is a rare genetic disorder characterized by features reminiscent of marked premature ageing^{1,2}. Here, we present evidence of mutations in lamin A (*LMNA*) as the cause of this disorder. The HGPS gene was initially localized to chromosome 1q by observing two cases of uniparental isodisomy of 1q—the inheritance of both copies of this material from one parent—and one case with a 6-megabase paternal interstitial deletion. Sequencing of *LMNA*, located in this interval and previously implicated in several other heritable disorders^{3,4}, revealed that 18 out of 20 classical cases of HGPS harboured an identical *de novo* (that is, newly arisen and not inherited) single-base substitution, G608G(GGC > GGT), within exon 11. One additional case was identified with a different substitution within the same codon. Both of these mutations result in activation of a cryptic splice site within exon 11, resulting in production of a protein product that deletes 50 amino acids near the carboxy terminus. Immunofluorescence of HGPS fibroblasts with antibodies directed against lamin A revealed that many cells show visible abnormalities of the nuclear membrane. The discovery of the molecular basis of this disease may shed light on the general phenomenon of human ageing.

Premature senescence is the hallmark of HGPS (Online Mendelian

Inheritance in Man (OMIM) 176670), both clinically and in cell culture^{1,2}. Children affected with HGPS typically appear normal at birth, but within a year the characteristic features of failure to thrive, delayed dentition, alopecia and sclerodermatous skin changes begin to appear. Death occurs on average at age 13, and at least 90% of the patients die from progressive atherosclerosis of the coronary and cerebrovascular arteries². Although the prevailing hypothesis of the inheritance of HGPS has been sporadic autosomal dominant^{5,6}, the possibility of autosomal recessive inheritance^{7,8} led us to conduct a genome-wide scan searching for evidence of homozygosity⁹. A whole-genome scan including 403 polymorphic microsatellite markers with an average spacing of 9.2 cM was performed on 12 DNA samples derived from individuals considered by one of us (W.T.B.) to represent classical HGPS. Although no evidence of homozygosity was identified in the overall sample set, two HGPS samples were found to have uniparental isodisomy (UPD) of chromosome 1q (Fig. 1a). Spectral karyotyping and G-banding of one of the two UPD cases showed a normal karyotype (data not shown).

An earlier report¹⁰ described an abnormal karyotype in a monozygotic twin with HGPS. That report described a mosaic cell population in which 70% of the cells contained a balanced inverted insertion (46,XY, inv ins (1;1)(q32;q44q23)), whereas the rest of the cells had an apparently normal karyotype. We obtained a fibroblast culture from the same individual (sample identification C8803), as well as his parents. Unexpectedly, genotyping of microsatellite markers identified a roughly 6-megabase (Mb) interval where all tested paternal alleles were completely missing (Fig. 1b). We confirmed that this deletion was also present in the cells that had an apparently normal karyotype, using fluorescent *in situ* hybridization (FISH) with bacterial artificial chromosomes (BACs) that map throughout this interval (Fig. 1c). Putting all of this information together with genotypes from a total of 44 additional microsatellite markers, we determined that the HGPS gene must lie in an interval of 4.82 Mb on proximal chromosome 1q (Fig. 1d).

The candidate interval contains roughly 80 known genes. Our attention was immediately drawn to one of them, the *LMNA* gene that encodes two protein products (lamin A and lamin C), representing major constituents of the inner nuclear membrane lamina. Mutations in *LMNA* have been found to be the cause of six different recessive and dominant disorders, including Emery–Dreifuss muscular dystrophy type 2, a form of dilated cardiomyopathy, the Dunnigan type of familial partial lipodystrophy, limb girdle muscular dystrophy type 1B, Charcot–Marie–Tooth disorder type 2B1, and mandibuloacral dysplasia (see reviews of laminopathies in refs 3, 4). The gene contains 12 exons and covers about 25 kilobases of genomic DNA. Lamin A is coded by exons 1–12 and lamin C by exons 1–10. A splice site within exon 10, located just upstream of the stop codon for lamin C, splices together with exons 11 and 12 to code for lamin A^{11,12}.

Polymerase chain reaction (PCR) amplification of all of the exons of the *LMNA* gene (including exon–intron boundaries), followed by direct sequencing, was carried out in 23 samples from patients with classical HGPS. In 18 of these samples a heterozygous base substitution (G608G(GGC > GGT)) within exon 11 of the *LMNA* gene was identified (Fig. 2a). In HGPS sample AG10801, a different heterozygous base substitution was identified within the same codon (G608S(GGC > AGC)) (Fig. 2a). In HGPS sample AG10677, we identified a heterozygous base substitution within exon 2 (E145K(GAG > AAG)). In the eight cases where DNA from both parents was available, the G608G mutation was absent in the parents, confirming that these are *de novo* mutations. Similarly, the G608S and E145K mutations were not found in parents of AG10801 or AG10677, respectively. Thus, of the 23 classical HGPS cases studied, there were only three in which no *LMNA* mutations were found (Table 1): the two UPD cases (AG10578 and HGADFN005) and the sample with the 6-Mb paternal deletion (C8803).

The *de novo* recurrence of the same exact point mutation in 18 out of 20 cases of classical HGPS is a surprising finding, but is not without precedent. The common HGPS mutation is a C to T in the context of a CpG dinucleotide, which represents the most mutable base in the vertebrate genome, as a methylated C can be deaminated to T and miscopied. In the E145K case and in four out of four G608G HGPS cases where parental DNA was available and silent polymorphisms in exon 10 (H566H) or exon 7 (D446D) were informative, we were able to determine that the new mutation was paternal in origin (data not shown). A very similar phenomenon occurs in achondroplasia¹³, where nearly all sporadic cases are due to paternal CpG to TpG mutations in the FGFR3 gene, resulting in an apparent gain-of-function mutation.

The most common mutation, G608G(GGC > GGT), is a silent substitution. The second mutation in the same codon,

G608S(GGC > AGC), results in a conservative substitution of serine for glycine. How is it possible that these bland-appearing *de novo* mutations could cause HGPS? Inspection of the normal sequence surrounding codon 608 reveals that both of the observed HGPS mutations improve the match to a consensus splice donor (Fig. 2b), suggesting that they might activate a cryptic splice site. To test this hypothesis, PCR with reverse transcription (RT) was performed using a forward primer spanning the junction of exons 7 and 8, and a reverse primer in exon 12. In RNA from unaffected individuals, the expected product appears (Fig. 2c). In RNA samples from cell lines harbouring HGPS mutations, an additional smaller RT-PCR product appears. Sequence of these fragments shows that 150 nucleotides within exon 11 are missing. As the reading frame is maintained, this abnormal transcript would be expected to code for a protein with an internal deletion of 50 amino acids near the C

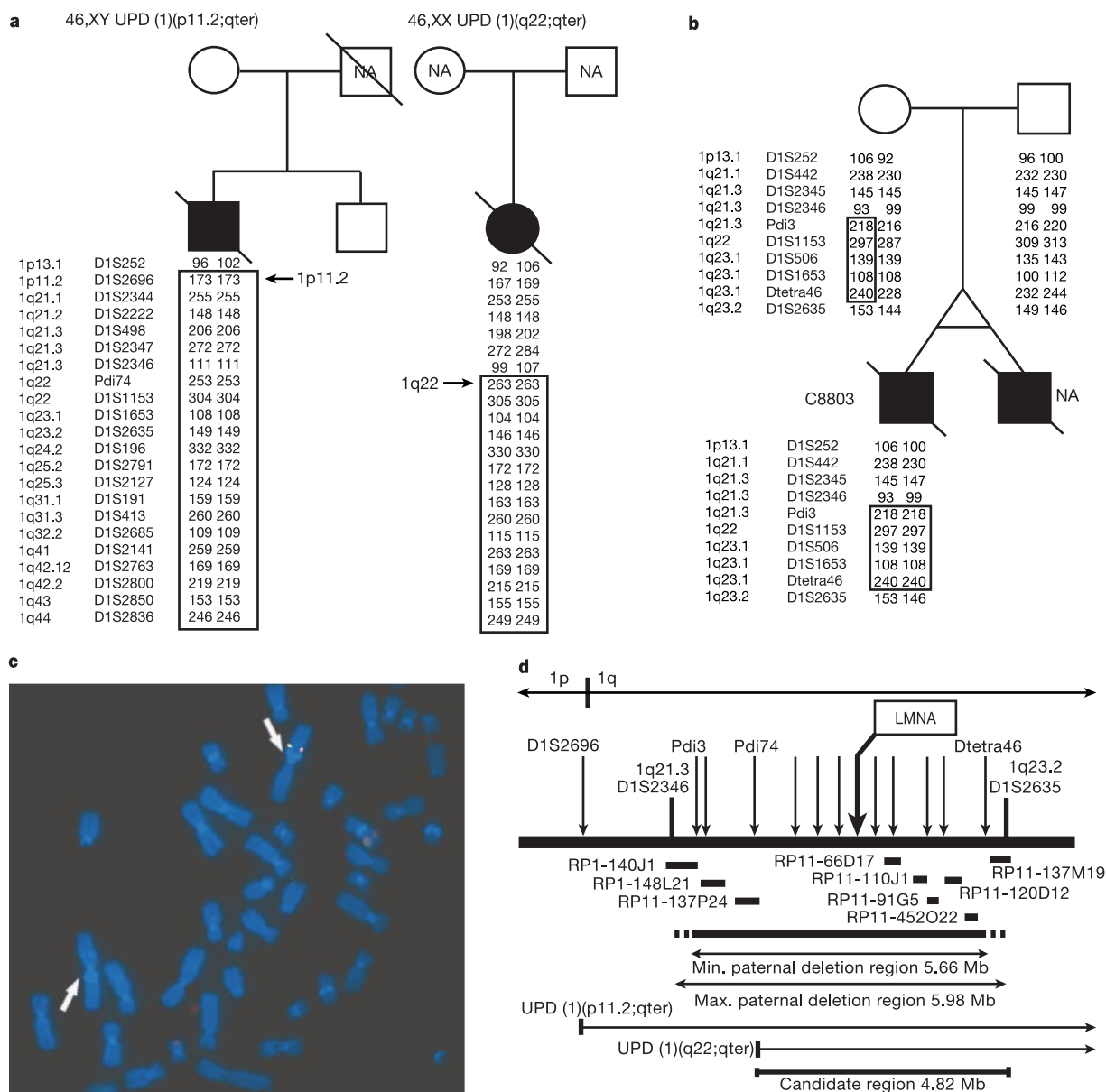


Figure 1 Localization of the HGPS gene to a 4.82-Mb interval on chromosome 1q. **a**, Uniparental isodisomy of chromosome 1q in fibroblasts from two patients with HGPS (filled symbols). A subset of markers and their genotypes are shown. NA, not available. **b**, The boxed interval is the region in proband C8803 that has been inherited exclusively from the mother. **c**, FISH analysis of a metaphase spread from C8803 fibroblasts,

hybridized with BAC probes RP11-110J1 (green) and RP11-91G5 (red). **d**, Summary map of the candidate region. Microsatellite markers are indicated with arrows; horizontal bars indicate BAC probes that were used for FISH on sample C8803. *LMNA* is one of the approximately 80 known genes in the 4.82-Mb candidate interval.

terminus of lamin A. Lamin C would be unaffected. Cloning and sequencing of the normal full-length fragment obtained from RT-PCR on RNA extracted from primary fibroblasts containing the more common codon 608 mutation revealed that 7 out of 23 clones carry the mutant sequence. Thus, activation of the cryptic

splice site within exon 11 is not complete. To determine whether the mutant messenger RNA is actually translated into protein, a western blot was performed using a monoclonal antibody against lamin A/C (Fig. 2d). In addition to the normal bands for lamin A and lamin C, an additional band is present in four of the lanes corresponding to samples from classical HGPS cases, but not in their parents. The abnormal band is not visible in the lane that contains the protein sample from classical HGPS patient AG11498 (which carries G608G(GGC > GGT)), but this is probably due to the very small amount of lamin A being expressed in this particular fibroblast line.

Although the major cause of HGPS appears to be the creation of an abnormal splice donor in exon 11, the finding of a *de novo* point mutation in exon 2 in a single patient (AG10677) is of interest. In retrospect, this patient¹⁴ had atypical clinical features (including persistence of coarse hair over the head, ample subcutaneous tissue over the arms and legs, and severe strokes beginning at age 4) that may subtly distinguish this phenotype from classical HGPS.

The two UPD cases present an interesting dilemma: if the *LMNA* gene sequence is normal in both cases, why do they have HGPS? The possibility of imprinting must be considered; however, previous cases of both paternal and maternal complete isodisomy of chromosome 1 do not support the presence of any imprinted loci on this chromosome^{15,16}. In the case where DNA samples were available from the mother and sibling (Fig. 1a), we can conclude that this phenomenon resulted in a chromosome that has partly paternal and partly maternal alleles. This must have arisen by a post-fertilization event, probably a mitotic crossover between homologues. We postulate that these cases actually represent 'somatic rescue' events of the premature senescence phenotype of HGPS. Under this hypothesis, the individuals from whom these fibroblasts were derived originally harboured typical codon 608 *LMNA* mutations. Perhaps as an *in vivo* event, or perhaps as an *in vitro* event in the fibroblast culture, a mitotic crossover occurred, generating a cell with segmental UPD that had duplicated the wild-type allele of *LMNA* and lost the HGPS mutation. Such a cell would probably then have a growth advantage over its neighbours. Proof of this hypothesis would require access to multiple tissues of the deceased UPD patients, which unfortunately are not available.

We also did not identify any mutation in the *LMNA* gene in the patient with the 6-Mb deletion (Fig. 1b). We believe that this might also be due to a somatic rescue event—specifically, we hypothesize that this patient was originally heterozygous for a codon 608 *LMNA* mutation, but in this instance the 'rescue' involved an internal deletion of 6 Mb containing the mutant allele, associated with a more complex mosaic rearrangement of chromosome 1. A recent study reported on a case of classical HGPS in which an apparent interstitial deletion of chromosome 1q23 was seen¹⁷. We obtained cells and DNA from this patient, and to our surprise the typical heterozygous G608G mutation is present (data not shown). Furthermore, we have not been able to confirm the presence of an interstitial deletion by high-resolution chromosome banding or by FISH with BACs spanning the 1q23–1q24 region. This may be another example of somatic rescue, involving an interstitial deletion in the clone of cells analysed in the original report, but not present in other samples from the same patient.

Extensive prior study of the biochemical function of lamin A suggests a possible mechanism for disease causation. Lamin A is normally synthesized as a precursor molecule (prolamin A). At the C terminus is a CAAX-box motif that is subject to farnesylation. After that, an internal proteolytic cleavage occurs, removing the last 18 coding amino acids^{18–20}, to generate mature lamin A. We predict that the codon 11 HGPS mutations and consequent abnormal splicing would produce a prolamin A that still retains the CAAX box, but is missing the site for endoproteolytic cleavage. There is also evidence that cell-cycle-dependent phosphorylation of lamin A is important for its normal function, and at least one site for phosphorylation (Ser 625) is deleted in the abnormal HGPS

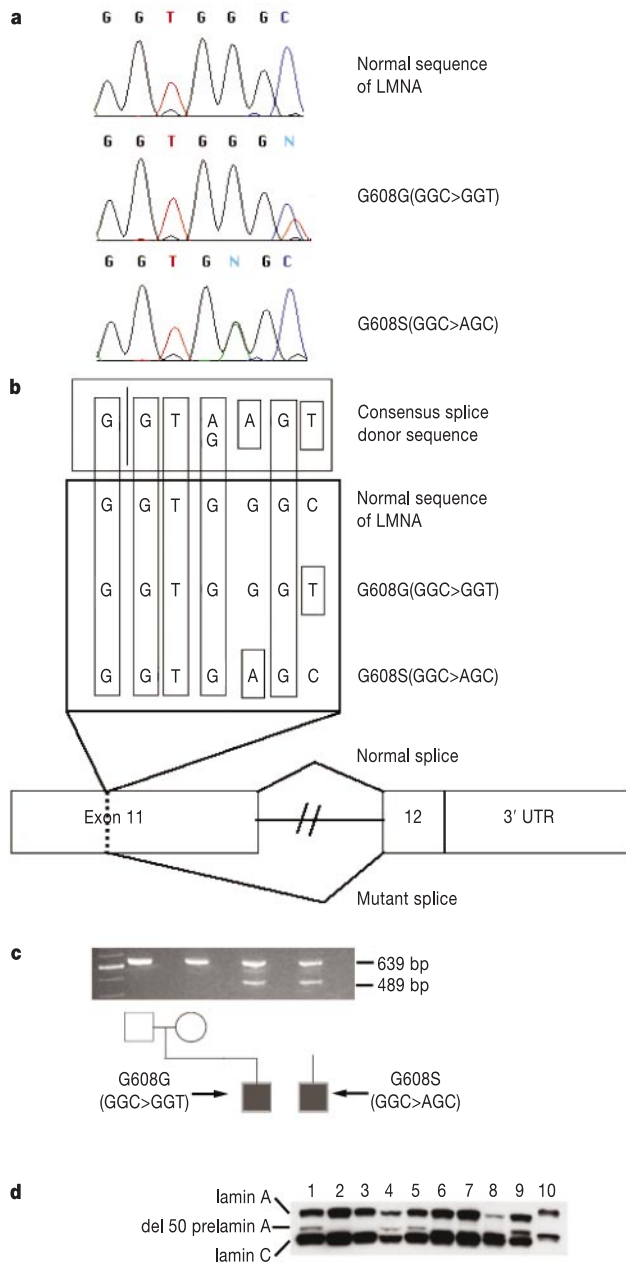


Figure 2 Point mutations in exon 11 of *LMNA* cause HGPS. **a**, Sequence traces from a normal control and two HGPS patients. **b**, Hypothesis for activation of a cryptic splice donor site in exon 11. **c**, Demonstration of the abnormal splice product using RT-PCR, showing an abnormal product of 489 bp in the two HGPS probands due to activation of the cryptic splice site. Alternative lanes to the right lack reverse transcriptase. **d**, Western blot using a monoclonal antibody against lamin A/C. Lanes 1, 5, 8 and 9 are from AG03506, AG03344, AG11498 and AG01972; they all carry G608G(GGC > GGT). Lane 4 is from AG10801, carrying G608S(GGC > AGC); lanes 2 and 3 are from parents of AG03506; lanes 6 and 7 are from father of AG03259 and mother of AG06917, respectively. Lanes 1–5 are from lymphoblastoid cell lines; lanes 6–9 are from fibroblasts. A protein sample from HeLa cells is in lane 10.

Table 1 *LMNA* mutations in samples from classical HGPS

Sample ID (classical HGPS)	Codon 608 nucleotide sequence	Mutation	Comment	Codon 608 nucleotide sequence		
				Mother	Father	Sibling(s)
AG01972	GGC/T	G608G	—	NA	NA	NA
AG06297	GGC/T	G608G	—	NA	NA	NA
AG11498	GGC/T	G608G	—	NA	NA	NA
AG11513	GGC/T	G608G	—	NA	NA	NA
AG03506	GGC/T	G608G	—	Normal	Normal	Normal
AG03344	GGC/T	G608G	—	Normal	Normal	Normal
AG03259	GGC/T	G608G	—	Normal	Normal	Normal
AG06917	GGC/T	G608G	—	Normal	Normal	NA
AG10579	GGC/T	G608G	—	NA	NA	NA
AG10587	GGC/T	G608G	—	Normal	NA	Normal
HGADFN001	GGC/T	G608G	—	NA	NA	NA
HGADFN003	GGC/T	G608G	—	NA	NA	NA
HGALBV009	GGC/T	G608G	—	Normal	Normal	NA
HGALBV011	GGC/T	G608G	—	Normal	Normal	NA
HGALBV057	GGC/T	G608G	—	Normal	Normal	NA
HGADFN008	GGC/T	G608G	—	NA	NA	NA
HGADFN014	GGC/T	G608G	—	NA	NA	NA
HGALBV071	GGC/T	G608G	—	NA	NA	NA
AG10801	A/GGC	G608S	—	Normal	Normal	NA
AG10578	GGC	—	UPD	Normal	NA	Normal
HGADFN005	GGC	—	UPD	NA	NA	NA
C8803	GGC	—	Deletion	Normal	Normal	NA
AG10677	GGC	E145K	—	Normal*	Normal*	NA

Additional sequence variants, presumed to be polymorphisms, were identified in exon 3 (L240L(CTG > CTA)), intron 4 (IVS4 + 61C > T), exon 5 (A287A(GCT > GCC)), exon 7 (D446D(GAT > GAC)), intron 8 (IVS8 – 41C > T) and exon 10 (H566H(CAC > CAT)). The variants in exons 5, 7 and 10 have been reported previously^{4,28}. NA, not available.
*Normal at codon 145.

protein²¹. As lamin A forms a multiprotein complex within the inner nuclear membrane, this incompletely processed prelamins A might well act as a dominant negative.
Notably, defective prelamins A processing has recently been identified in a mouse knockout of the Zmpste24 metalloproteinase^{22,23}. Zmpste24 is believed to be involved in proteolytic processing of prelamins A, and may represent the actual endoprotease. The homozygote knockout of Zmpste24 has a phenotype resembling the clinical features observed in HGPS patients, including growth retardation, premature death (20 weeks) from cardiac dysfunction, and alopecia. However, additional features such as pronounced osteoporosis are also present²³. Immunofluorescence experiments on cells from these animals show elongated and irregular nuclei with herniation of the nuclear membrane. A mouse knockout of the *Lmna* gene has also been reported previously²⁴, showing severe

postnatal growth retardation, muscular dystrophy and similar nuclear abnormalities.
To pursue the possibility of structural nuclear membrane abnormalities in HGPS, immunofluorescence studies with two different monoclonal antibodies against lamin A/C were performed on primary fibroblasts from five classical HGPS cases with the common G608G mutation, as well as unaffected parents. Representative data from an affected patient (AG11498) and an unaffected parent (AG06299) are shown in Fig. 3. In 48% of the HGPS cells we noted an irregular shape of the nuclear envelope (Fig. 3e–h). Cells from the unaffected control (Fig. 3a–d) showed significantly fewer cells of this phenotype (<6%). To be certain that this result was not an artefact of secondary differences in the status of the HGPS and control fibroblast cultures, cells originating from the same flasks as the cells used for the immunofluorescence studies

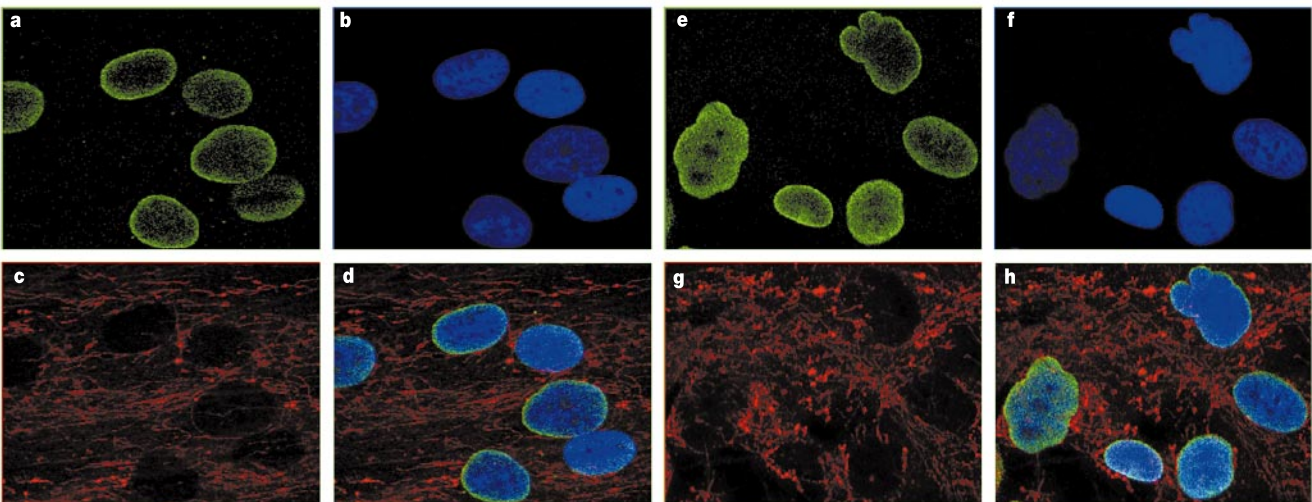


Figure 3 Immunofluorescence on primary dermal fibroblasts from an unaffected mother and a child with classical HGPS, using antibody JOL2 against lamin A/C. Identical results were obtained with antibody XB10 (data not shown). **a–d**, Results from an unaffected mother, AG06299. **e–h**, Results from a classical HGPS patient, AG11498. **a, e**, Antibody against lamin A/C. **b, f**, DAPI staining showing location of the nuclei. **c, g**, Antibody staining mitochondria, showing distribution of the cytosol. **d, h**, Merged image of **a–c** (**d**) and **e–g** (**h**).

were monitored for differences in cell cycle (by fluorescent-activated cell analysis) and degree of apoptosis. No significant differences between the cells derived from the classical HGPS patients and the unaffected parents were noted.

HGPS can now be added to the long list of human genetic disorders shown to be due to mutations in the *LMNA* gene^{3,4}. As most cases of HGPS appear to have a *de novo* mutation in the same codon, molecular diagnostics should be almost immediately feasible. This will be particularly useful in making the diagnosis in a young child before the full clinical phenotype has appeared. Molecular diagnostic methods may also provide reassurance in the prenatal arena, where the possibility of parental somatic mosaicism and recurrence of disease in future pregnancies can now be addressed. Finally, the discovery of the molecular basis of HGPS suggests a possible role for *LMNA* in aspects of the normal ageing process. It will be important to look for common variants in this gene that might show association with exceptional longevity, and perhaps also to explore whether somatic mutations in *LMNA*, accumulated over a lifetime, have some role in senescence. □

Methods

Subjects and DNA/RNA preparation

This study was approved through the NIH Human Subjects review process. Primary dermal fibroblast cell cultures and EBV-transformed lymphoblastoid cell lines from individuals diagnosed as classical HGPS, and their first-degree relatives (when available), were obtained from the Aging Repository of the Coriell Cell Repository (CCR), and the Progeria Research Foundation Cell and Tissue Bank. DNA was prepared using the Puregene DNA isolation kit (Gentra Systems). The genome scan for homozygosity included samples derived from 12 classical HGPS patients (samples AG01972, AG03259, AG03344, AG03506, AG06297, AG06917, AG10578, AG10579, AG10587, AG10801, AG11498 and AG11513) and 16 unaffected first-degree relatives (samples AG03258, AG03260, AG03262, AG03263, AG03343, AG03342, AG03345, AG03346, AG03504, AG03505, AG03507, AG03508, AG06298, AG06299, AG10585 and AG10588). Additional samples used in this study were samples from classical HGPS (samples AG10677, HGADFN001, HGADFN003, HGALBV009, HGALBV011, HGALBV057, HGADFN005, HGADFN008, HGADFN014, HGALBV071 and C8803 (also known as AG10548 in the CCR)) and samples derived from their unaffected first-degree relatives (samples HGMLBV010, HGFLBV021, HGMLBV023, HGFLBV031, HGMLBV066, HGFLBV067, HGMLBV013, HGFLBV050, HGMLBV058, HGSLBV059, HGMLBV078 HGFLBV079, HGFLBV082 and HGMLBV081). Total RNA was extracted from cells with TRIzol reagent (Invitrogen).

Genotyping

The whole-genome scan included 403 highly polymorphic microsatellite markers with an average spacing of 9.2 cM and average heterozygosity of about 0.8 (E. Gillanders, M. Jones, D. Freas-Lutz, R. Sood and J. Trent, manuscript in preparation). Pedigree checking was performed using PedCheck (J. O'Connell, <http://watson.hgen.pitt.edu/register/docs/pedcheck.html>) and any identified genotype errors were removed. We carried out homozygosity mapping assuming various degrees of inbreeding for the HGPS cases²⁵. Additional microsatellite repeats on chromosome 1q were identified using the Sputnik program (C. Abajian, <http://rast.abajian.com/sputnik/>) (see Supplementary Table 1) and were used to further investigate the UPD cases and the paternal deletion region in C8803. Microsatellite markers were analysed using a 3100 genetic analyser (PE Biosystems). Genotypes were analysed using GeneScan 3.7 and Genotyper 2.5 Software (PE Biosystems).

FISH analysis

Single- and two-colour FISH was performed on metaphases from sample C8803 following previously published procedures²⁶, using a subset of BACs in the region of the paternal deletion. BACs used as probes for the FISH were RP1-140J1, RP1-148L21, RP1-178F15, RP11-137P24, RP11-66D17, RP11-110J1, RP11-91G5, RP11-120D12, RP11-101J8, RP11-81N17, RP11-144L1, RP11-317F9, RP11-452O22 and RP11-137M19.

Mutation analysis of *LMNA*

Direct sequencing of *LMNA* was performed primarily using previously described primer sequences for the *LMNA* exons 1–12 (ref. 27). Sequencing reactions were performed at quarter strength reaction volumes with the Big Dye Terminator chemistry kit (Applied Biosystems), and electrophoresed on an ABI 3700 DNA Analyser (Applied Biosystems). Multiple sequence alignment was performed with Sequencher (Genecodes Inc.). Attempts were made to sequence all PCR products in both directions, but approximately 7% of exons failed to yield readable sequence.

RT-PCR on exon 11

For all RNA samples, 20 µg of total RNA was treated with RQ1 RNase-Free DNase according to manufacturers recommendations (Promega). A total of 800 ng of DNase-treated total RNA was used for first-strand complementary DNA synthesis with random hexamers (Superscript, Invitrogen). Control samples without reverse transcriptase were

processed at the same time for each sample. PCR primers for the lamin A/C gene were designed in exon 7/8 and exon 12. Primer sequences were 5'-GCAACAAGTCCAATGA GGACCA-3' and 5'-GTCCCAGATTACATGATGC-3'. PCR fragments were gel-purified or cloned (TOPO TA-cloning kit, Invitrogen) and sequenced. PCR with GAPDH-specific primers was performed on all samples as control.

Western analysis

Whole cells (1×10^6) were collected and washed twice in PBS, the pellets were resuspended in RIPA buffer and included a cocktail of proteinase inhibitors (Roche). A total of 20 µg of protein was loaded and electrophoresed on an 8% Tris-Glycine mini gel (Invitrogen). Primary monoclonal lamin A/C antibody was clone JOL2 (Chemicon International).

Immunofluorescence

Cultivated fibroblasts were fixed (3.2% paraformaldehyde), permeabilized (1% NP-40), blocked (0.1% Brij58 and 5% serum) and labelled for lamin A/C (monoclonal antibodies JOL2, Chemicon International, and clone XB10, CRP Inc.) mitochondria (HMS-0100, Immunovision Inc.) and DNA (using 4,6-diamidino-2-phenylindole (DAPI) dihydrochloride mixed in at $1 \mu\text{g ml}^{-1}$ with the second antibody). Analysis was done by confocal microscopy using a Biorad 1024 and Leica SP2 systems.

Cell cycle and apoptosis

Cells were collected and washed in PBS one day after the immunofluorescence experiments. Duplicate experiments were performed on each cell culture. A total of 5×10^5 cells were resuspended in 0.5 ml of NuCycl propidium iodide (NuCycl PI kit, Alpha) and processed as recommended by the manufacturer. The total DNA content was measured by DNA flow cytometry. Cells were also assayed for viability using annexin V-fluorescein isothiocyanate and propidium iodide according to standardized procedures (BD Biosciences).

Received 24 January; accepted 1 April 2003; doi:10.1038/nature01629.

Published online 25 April 2003.

- DeBusk, F. L. The Hutchinson-Gilford progeria syndrome. *J. Pediatr.* **80**, 697–724 (1972).
- Baker, P. B., Baba, N. & Boesel, C. P. Cardiovascular abnormalities in progeria. Case report and review of the literature. *Arch. Pathol. Lab. Med.* **105**, 384–386 (1981).
- Burke, B. & Stewart, C. L. Life at the edge: the nuclear envelope and human disease. *Nature Rev.* **3**, 575–585 (2002).
- Genschel, J. & Schmidt, H.-J. Mutations in the *LMNA* gene encoding lamin A/C. *Hum. Mutat.* **16**, 451–459 (2000).
- Brown, W. T. Human mutations affecting aging—a review. *Mech. Aging Dev.* **9**, 325–336 (1979).
- Brown, W. T., Zebrower, M. & Kieras, F. J. in *Genetic Effects on Aging II* (ed. Harrison, D. E.) 521–542 (Telford, Caldwell, New Jersey, 1990).
- Trevas-Maciell, A. Evidence for autosomal recessive inheritance of progeria (Hutchinson Gilford). *Am. J. Med. Genet.* **31**, 483–487 (1988).
- Khalifa, M. M. Hutchinson-Gilford progeria syndrome: report of a Libyan family and evidence of autosomal recessive inheritance. *Clin. Genet.* **35**, 125–132 (1989).
- Lander, E. S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567–1570 (1987).
- Brown, W. T. Progeria: a human disease model of accelerated aging. *Am. J. Clin. Nutr.* **55**, 1222S–1224S (1992).
- Fisher, D. Z., Chaudhary, N. & Blobel, G. cDNA sequencing of nuclear lamins A and C reveals primary and secondary structural homology to intermediate filament proteins. *Proc. Natl Acad. Sci. USA* **83**, 6450–6454 (1986).
- Lin, F. & Worman, H. J. Structural organization of the human gene encoding nuclear lamin A and nuclear lamin C. *J. Biol. Chem.* **268**, 16321–16326 (1993).
- Shiang, R. *et al.* Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell* **78**, 335–342 (1994).
- Smith, A. S., Witznitzer, M., Karaman, B. A., Horwitz, S. J. & Lanzieri, C. F. MRA detection of vascular occlusion in a child with progeria. *Am. J. Neuroradiol.* **14**, 441–443 (1993).
- Pulkkinen, L., Bullrich, F., Czarnecki, P., Weiss, L. & Uitto, J. Maternal uniparental disomy of chromosome 1 with reduction to homozygosity of the LAMB3 locus in a patient with Herlitz junctional epidermolysis bullosa. *Am. J. Hum. Genet.* **61**, 611–619 (1997).
- Gelb, B. D. *et al.* Paternal uniparental disomy for chromosome 1 revealed by molecular analysis of a patient with pycnodysostosis. *Am. J. Hum. Genet.* **62**, 848–854 (1998).
- Delgado-Luengo, W. *et al.* Del(1)(q23) in a patient with Hutchinson-Gilford progeria. *Am. J. Med. Genet.* **113**, 298–301 (2002).
- Lutz, R. J., Trujillo, M. A., Denham, K. S., Wenger, L. & Sinensky, M. Nucleoplasmic localization of prelamin A: implications for prenylation-dependent lamin A assembly into the nuclear lamina. *Proc. Natl Acad. Sci. USA* **89**, 3000–3004 (1992).
- Sinensky, M. *et al.* The processing pathway of prelamin A. *J. Cell Sci.* **107**, 61–67 (1994).
- Hennekes, H. & Nigg, E. A. The role of isoprenylation in membrane attachment of nuclear lamins, a single point mutation prevents the proteolytic cleavage of the lamin A precursor and confers membrane binding properties. *J. Cell Sci.* **107**, 1019–1029 (1994).
- Eggert, M. *et al.* Identification of novel phosphorylation sites in murine A-type lamins. *Eur. J. Biochem.* **213**, 659–671 (1993).
- Pendas, A. M. *et al.* Defective prelamin A processing and muscular and adipocyte alterations in Zmpste24 metalloproteinase-deficient mice. *Nature Genet.* **31**, 94–99 (2002).
- Bergo, M. O. *et al.* Zmpste24 deficiency in mice causes spontaneous bone fractures, muscle weakness, and a prelamin A processing defect. *Proc. Natl Acad. Sci. USA* **99**, 13049–13054 (2002).
- Sullivan, T. *et al.* Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J. Cell Biol.* **147**, 913–920 (1999).
- Smith, C. The detection of linkage in human genetics. *J. R. Stat. Soc. B* **15**, 153–184 (1953).
- Casper, A. M., Nghiem, P., Arlt, M. F. & Glover, T. W. ATR regulates fragile site stability. *Cell* **111**, 779–789 (2002).

27. De Sandre-Giovannoli, A. *et al.* Homozygous defects in *LMNA*, encoding lamin A/C nuclear-envelope proteins, cause autosomal recessive axonal neuropathy in human (Charcot-Marie-Tooth disorder type 2) and mouse. *Am. J. Hum. Genet.* **70**, 726–736 (2002).
28. Speckman, R. A. *et al.* Mutational and haplotype analyses of families with familial partial lipodystrophy (Dunnigan variety) reveal recurrent missense mutations in the globular C-terminal domain of lamin A/C. *Am. J. Hum. Genet.* **66**, 1192–1198 (2000); erratum **67**, 775 (2000).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to all of the families with HGPS who donated blood and skin samples, without which this study would not have been possible. We would also like to express our gratitude to the entire Progeria Research Consortium, especially B. Toole, for their encouragement during the course of this work. We also thank E. Gillanders for providing the linkage marker set; P. Hollstein for assistance with genotyping; P. Chines for the identification of additional microsatellite markers on chromosome 1q; and M. Kirby for assistance with the flow cytometry. Support from the Progeria Research Foundation (L.B.G., M.W.G. and T.W.G.) and the Bedminster Foundation (W.T.B.) is gratefully acknowledged.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.S.C. (fc23a@nih.gov).

A progeroid syndrome in mice is caused by defects in A-type lamins

Leslie C. Mounkes, Serguei Kozlov, Lidia Hernandez, Teresa Sullivan & Colin L. Stewart

Cancer and Developmental Biology Laboratory, National Cancer Institute, Frederick, Maryland 21702, USA

Numerous studies of the underlying causes of ageing have been attempted by examining diseases associated with premature ageing, such as Werner's syndrome and Hutchinson–Gilford progeria syndrome (HGPS). HGPS is a rare genetic disorder resulting in phenotypes suggestive of accelerated ageing, including shortened stature, craniofacial disproportion, very thin skin, alopecia and osteoporosis, with death in the early teens predominantly due to atherosclerosis¹. However, recent reports suggest that developmental abnormalities may also be important in HGPS^{1,2}. Here we describe the derivation of mice carrying an autosomal recessive mutation in the lamin A gene (*Lmna*) encoding A-type lamins, major components of the nuclear lamina³. Homozygous mice display defects consistent with HGPS, including a marked reduction in growth rate and death by 4 weeks of age. Pathologies in bone, muscle and skin are also consistent with progeria. The *Lmna* mutation resulted in nuclear morphology defects and decreased lifespan of homozygous fibroblasts, suggesting premature cell death. Here we present a mouse model for progeria that may elucidate mechanisms of ageing and development in certain tissue types, especially those developing from the mesenchymal cell lineage.

The *Lmna* gene encodes the A-type lamins, intermediate filament family members that make up the nuclear lamina, a fibrous network underlying the nuclear envelope³. Alternative splicing of *Lmna* transcripts gives rise to four proteins: two major products called lamin A and lamin C, and two minor products, lamin AΔ10 and lamin C2, the latter being specific to the testis. The A-type lamins interact with B-type lamins, which are encoded by the *Lmnb1* and *Lmnb2* genes, in a largely uncharacterized assembly process to form the nuclear lamina. The A-type lamins are developmentally regulated, whereas at least one B-type lamin is expressed in every cell

type at all developmental stages. Embryonic stem cells, cells of the early embryo and haematopoietic stem cells do not express *Lmna*; however, derivatives of these cells do express *Lmna* during development, suggesting a role for *Lmna* in terminal differentiation⁴. In addition to controlling interphase nuclear morphology, the nuclear lamina also functions in selective retention of proteins in the inner nuclear membrane, chromatin organization, DNA replication and gene expression^{5,6}. A role for A-type lamins in regulating gene expression is further supported by interactions with a growing number of transcription factors, for example, Rb⁷ and GCL indirectly through emerin⁸.

The indispensable role of lamins in these diverse and fundamental processes may account for their association with a growing list of human diseases. Laminopathies can be divided loosely into two categories: those affecting striated muscle and those with phenotypes affecting adipose tissue and bone. Diseases of striated muscle caused by mutations in the *Lmna* gene are the autosomal dominant form of Emery–Dreifuss muscular dystrophy (AD-EDMD), limb girdle muscular dystrophy 1B (LGMD1B) and dilated cardiomyopathy with conduction system disease (DCM-CD). A single recessive mutation in *Lmna* has also been implicated in Charcot–Marie–Tooth syndrome type 2B1 (CMT2B1), a peripheral neuropathy with muscle weakness and wasting. Dunnigan's familial partial lipodystrophy (FPLD) and mandibuloacral dysplasia (MAD), disorders primarily resulting in loss and redistribution of white adipose tissue, are also associated with mutations in *Lmna*. Hyperlipidaemia, insulin resistance and diabetes are common in FPLD and MAD patients, and bone defects in MAD patients include craniofacial abnormalities, osteolysis of terminal digits and hypoplasia of clavicles⁹. Many clinical observations of MAD patients are also seen in progeria patients¹⁰, suggesting that the two disorders may be allelic. Although we do not understand why mutations in *Lmna* cause such a wide array of phenotypes, the continued discovery of phenotypic overlap between some of these diseases indicates that they may represent a spectrum of related disorders rather than separate diseases. The present description of progeroid symptoms associated with *Lmna*^{L530P/L530P} mutant

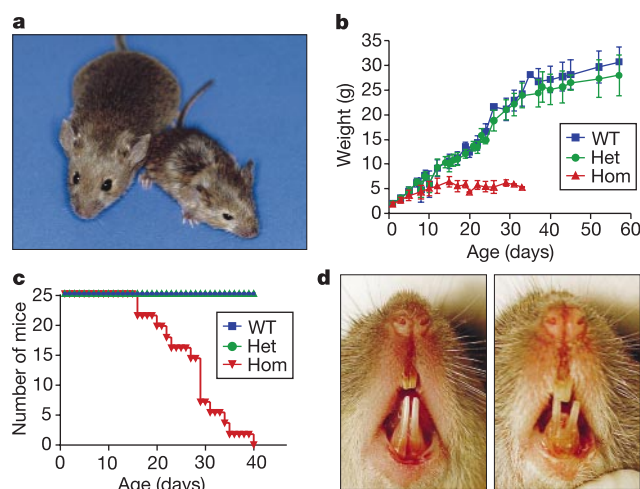


Figure 1 *Lmna*^{L530P/L530P} homozygous mice exhibit severely retarded growth and die early. **a**, The smaller *Lmna*^{L530P/L530P} mouse weighed 5.37 g; the littermate (21.88 g) was wild type. **b**, Mutants failed to thrive, showing marked growth reduction as early as 4 days after birth. Mice were weighed every 2–3 days for 58 days. *n* = 8 mutant (Hom, red), 10 wild-type (WT, blue) and 14 heterozygotes (Het, green) mice. **c**, Heterozygotes (green) can live as long as wild-type littermates (blue). Mutants (*n* = 25, red) died by 38 days. **d**, In 50% of mutant mice, we observed abnormal dentition in the lower jaw. A heterozygous littermate is on the left.

mice represents one of the most severe extremes in such a spectrum of laminopathies.

In an effort to create a mouse model for AD-EDMD, we introduced a nucleotide base change into the *Lmna* gene, directing a substitution of proline for leucine at residue 530 ($L530P$)¹¹—a mutation that causes AD-EDMD in humans. Although mice heterozygous for this point mutation do not show signs of muscular dystrophy and have been overtly normal up to 6 months of age, mice homozygous for the mutation ($Lmna^{L530P/L530P}$ mice) showed phenotypes markedly reminiscent of symptoms observed in progeria patients. Homozygous $Lmna^{L530P/L530P}$ mice are indistinguishable from their littermates at birth, but within 4–6 days develop severe growth retardation, dying within 4–5 weeks (Fig. 1a–c). Progeria patients are unremarkable at birth, but by 2 years of age develop severe growth retardation resulting in

shortened stature, and the mean age of death is 12–15 years. Homozygous $Lmna^{L530P/L530P}$ mice showed a slight waddling gait, suggesting immobility of joints. Similarly, progeria patients often adopt a 'horse-riding stance' with wide, shuffling gait due to joint stricture¹². Other progeroid features of $Lmna^{L530P/L530P}$ mice include micrognathia and abnormal dentition¹²—in approximately half of the mutants a gap was observed between the lower two incisors, which also appeared yellowed (Fig. 1d). Such tooth gaps and discoloration were not observed in age-matched wild-type or heterozygous mice. Mutant mice continue to feed and defecate until death, and gonadal fat pads are present, suggesting that insufficient food consumption did not cause the decreased growth rate. On suspension by their tails, homozygous mutant mice did not clasp their hind legs, suggesting a normal peripheral motor response. No obvious abnormalities in social behaviour or grooming were noted in the mutant mice. Mentation, IQ and personality are normal in progeria patients^{12,13}.

Pathology of the heart, skin, skeletal muscle and bone of 4-week-old homozygous $Lmna^{L530P/L530P}$ mice suggested developmental defects consistent with progeroid phenotypes (Fig. 2). The most obvious sign of premature ageing occurred in the skin of mutant mice (Fig. 2e, f), which showed an epidermal layer thickened by regions of hyperkeratosis, thinning of the underlying dermis, atrophy of the accompanying muscle, and a complete absence of the subcutaneous fat layer. Increased deposits of collagen, similar to scleroderma, were observed in skin sections stained with Masson's trichrome. A decreased density of hair follicles observed in mutant skin sections may correlate to the alopecia observed in progeria patients; reduced numbers of eccrine and sebaceous glands were also observed in mutants. Mild to moderate degeneration mixed with hypoplasia and/or atrophy of both heart and skeletal muscle

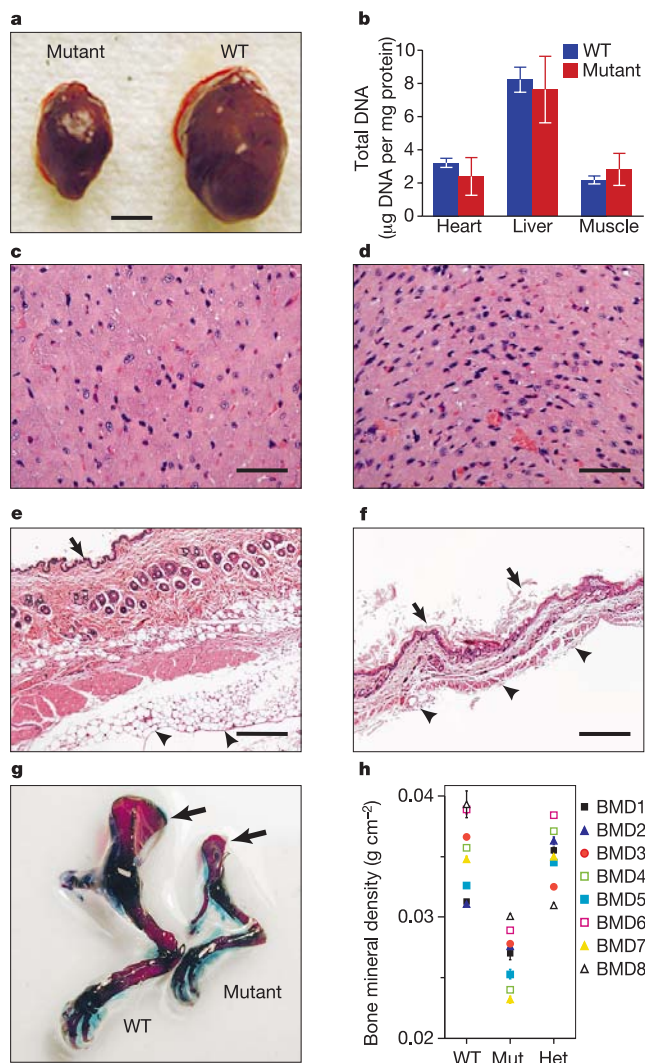


Figure 2 Tissue pathology in $Lmna^{L530P/L530P}$ homozygous mutant mice. **a**, The hearts of $Lmna^{L530P/L530P}$ mice (left) are smaller than age-matched wild-type hearts (right). Scale bar, 2 mm. **b**, Similar quantities of DNA were found in mutant and wild-type organs. **c**, **d**, Hearts of $Lmna^{L530P/L530P}$ mice (**d**) show smaller myocytes than hearts from wild-type mice (**c**). **e**, **f**, Skin from wild-type (**e**) and mutant (**f**) mice. Arrows indicate epidermis, with hyperkeratotic regions in the mutant. Arrowheads indicate subcutaneous fat, of which there is an almost complete absence in the mutant. **g**, Skeletal preparations of wild-type and mutant mice revealed malformed, thin scapulae in $Lmna^{L530P/L530P}$ mice. **h**, Bone mineral contents ($g\ cm^{-2}$) of wild-type, heterozygote and mutant mice are shown. $n = 8$ per genotype.

Table 1 Comparison of HGPS phenotypes in human and $Lmna^{L530P/L530P}$ mouse	
Human	Mouse
Severe growth retardation	Severe growth retardation
Short stature; failure to thrive	Short stature; failure to thrive
Mean death at 12–15 years	Death at 4–5 weeks
Craniofacial disproportion; micrognathia	Micrognathia
Abnormal dentition	Abnormal dentition
Very thin skin; loss subcutaneous fat	Loss of subcutaneous fat
Decreased eccrine, sebaceous glands	Decreased eccrine, sebaceous glands
Scleroderma	Increased collagen deposition in skin
Alopecia, onset at about 1 year	Decreased hair follicle density
Hyperkeratosis in some patients	Hyperkeratosis
Bone hypoplasia and resorption; osteoporosis	Decreased bone density; thin trabeculae
Hypoplasia/resorption of clavicles	Malformation of scapulae
Resorption of hip-girdle joints; shuffling gait	Waddling gait
Congestive heart failure; decrease in vascular smooth muscle	Heart pathology; subtle changes consistent with pulmonary hypertension
Incomplete sexual maturation	Hypogonadism
No consistent dyslipidaemia	Normal TTG and FFA; low cholesterol
Hypoplastic facial bones	Not determined
Thin diaphyses	Thinner femur; other diaphyses not studied
Osteolysis of terminal digits	None observed
Protruding ears; prominent eyes	Protruding ears
Myocardial fibrosis	Increased cardiac collagen and fibrocyte number
Atherosclerosis	No obvious defects in aorta, small vessels
Poor muscle development; atrophy	Poor muscle development and/or atrophy

Online Mendelian Inheritance in Man (OMIM) cites HGPS cases from two families in which consanguineous siblings were affected with progeroid syndrome, suggesting possible autosomal recessive inheritance. However, many HGPS patients arise from a non-consanguineous union, suggesting an autosomal dominant mode of inheritance. Both modes of inheritance may be possible. $Lmna^{L530P/L530P}$ mice are consistent with the autosomal recessive inheritance pattern. TTG, total triglycerides; FFA, free fatty acids.

also occurred in mutant animals. Mutant mice had significantly smaller hearts (Fig. 2a) with myocytes smaller than their wild-type counterparts (Fig. 2c, d). The amount of DNA in known weights of heart tissue was unchanged in mutants, suggesting that the smaller mutant hearts had the same number of cells as wild-type hearts (Fig. 2b). Counts of fibrocyte versus myocyte nuclei in mutant and wild-type hearts indicated that there were 2.6 fibrocytes per myocyte in wild-type hearts, but 4.1 fibrocytes per myocyte in mutant hearts. Although the total number of cells may be equivalent in mutant and wild-type hearts, there appeared to be fewer myocytes in mutant tissue with a compensatory increase in the fibrocyte population. Although not profound, a slight increase in extracellular collagen, indicative of degeneration, was noted in mutant hearts by Masson's trichrome staining (data not shown). Hypoplasia and/or atrophy of oesophageal muscles and myocardiocytes in homozygous *Lmna*^{L530P/L530P} mice suggested either incomplete development or loss of muscle mass in these tissues. Mild to moderate degeneration of other skeletal muscles, including pharyngeal, paravertebral, shoulder, hip, tibia, radius, femur, humerus and tongue, was also observed. However, dystrophic features such as centrally located nuclei and muscle fibres of varying diameter were not observed, ruling out muscular dystrophy, a prominent feature in *Lmna*-null mice¹⁴.

Bone sections from 4-week-old mutant mice showed a decreased number and size of trabeculae, and a decrease in cortex width of vertebrae and femur, consistent with osteoporosis. In 50% of mutant mice surveyed, at least one scapula appeared to be smaller, thinner and misshapen compared with age-matched controls

(Fig. 2g), suggesting incomplete development of the shoulder blade. Such defects in a quadruped where scapulae experience more perpendicular stress than the clavicle may correlate with clavicular malformation in MAD and progeria patients. The xiphisternum, formed of cartilage at the lower front midline of the rib cage, also appeared smaller in mutant animals in relation to overall ribcage size. No osteolysis of terminal digits was observed. Densitometry scanning of whole mice indicated a decrease in total bone mineral density of 25–30% in mutant animals compared with wild-type or heterozygous littermates (Fig. 2h). Taken together these results suggest either a premature loss of bone mass or the incomplete development of the skeleton in mice homozygous for *Lmna*^{L530P/L530P}, features prominent in patients with HGPS². Consistent with several case histories of progeria subjects^{12,15}, juvenile testes and ovaries were observed in *Lmna*^{L530P/L530P} mice, indicating incomplete gonadal development. Total necropsy of the animals revealed no incidence of tumours or gross organ deficiencies. A comparison of progeroid phenotypes and those of *Lmna*^{L530P/L530P} mice is summarized in Table 1.

At the cellular level, nuclear morphology is disrupted in primary mouse embryo fibroblasts (MEFs) from *Lmna*^{L530P/L530P} homozygous embryos, consistent with phenotypes observed in fibroblast cell lines from AD-EDMD, FPLD and DCM patients^{16,17}. Discontinuities of the nuclear envelope were seen in 58% of mutant cells (Fig. 3b), whereas only 20% of heterozygous and 5.6% of wild-type cells showed herniations of the nucleus. Areas in which the inner nuclear membrane and outer nuclear membrane appeared to lose contact with each other, resulting in the expansion of the perinuclear space and ballooning of chromatin into a bleb-like structure, may reflect a weakened nuclear lamina in cells expressing *Lmna*^{L530P/L530P}. Consistent with a defective nuclear lamina, decreased levels of lamin A were incorporated into the lamina as deduced by indirect immunofluorescence (Fig. 3a, b). Lamin C is completely absent from the nuclear envelope in mutant primary MEFs, and there is an increase in lamin-C-reactive peptides in the cytoplasm, an observation not seen in wild-type cells when stained with the same antibody (Fig. 3c, d). These results suggest that the mutant L530P form of lamin A is competent to assemble at least partially or temporarily in the nuclear lamina, but lamin C is not assembled into the nuclear envelope in *Lmna*^{L530P/L530P} cells. Similar defective lamin A and C incorporation into the lamina was observed in *Lmna*-null primary MEFs when a mutant complementary DNA encoding the L530P variant of *Lmna* was expressed¹⁸. Western analysis indicated that any A-type lamins translated are unstable, as greatly diminished levels of mutant protein were detected in homozygous mutant cell lines. Cells heterozygous for *Lmna*^{L530P/L530P} had about half the wild-type amount of lamin A and C protein. Consistent with low levels of protein, decreased levels of *Lmna* transcripts were detected by northern analysis in cells from homozygous embryos, suggesting instability of the mutant message, probably due to a splicing defect. Polymerase chain reaction with reverse transcription (RT-PCR) of the region surrounding exon 9, which encodes residue 530, suggests aberrant splicing between exons 9 and 10 in *Lmna*^{L530P/L530P} mice (see Supplementary Information for details).

To determine whether the mutant cells enter senescence earlier than normal as would be expected of prematurely ageing cells, we performed continuous passage assays¹⁹ (Fig. 3e). Wild-type and heterozygous cell lines entered a senescent state at passage 10 to 11; these cells remained viable and attached to the tissue culture surface but did not continue dividing during this period. In contrast, cells homozygous for *Lmna*^{L530P/L530P} stopped dividing and began to disappear from the cultures by passage 8, suggesting that they were dying rather than attaining replicative senescence. Short-term growth characteristics of adult muscle fibroblasts at low cell density were equivalent to wild-type and heterozygous cells at early passage (passage 2), but by passage 6, the mutant cells failed to double over a

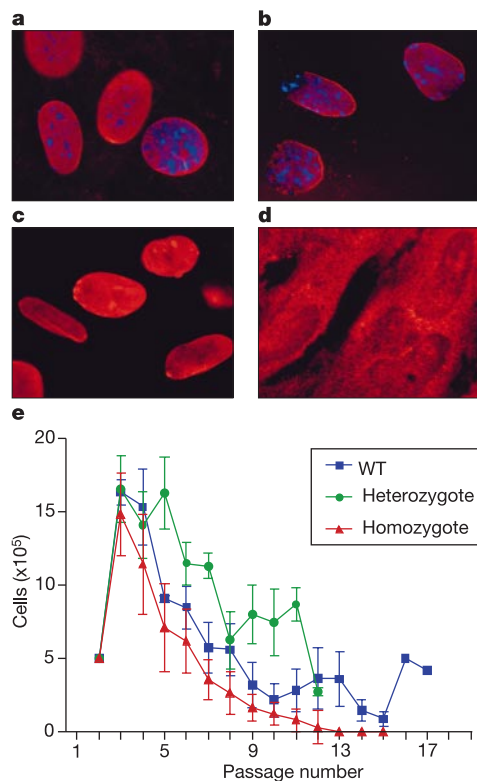


Figure 3 Homozygous *Lmna*^{L530P/L530P} fibroblasts showed nuclear envelope abnormalities and decreased lifespan. **a–d**, Indirect immunofluorescence for lamin A (**a** and **b**) or lamin C (**c**, **d**) showed localization of A-type lamins in wild-type (**a**, **c**) or homozygous mutant (**b**, **d**) primary MEFs. 4,6-Diamidino-2-phenylindole counterstaining is shown in blue (**a**, **b**). **e**, Continuous passage assays show decreased doubling capacity and lifespan of *Lmna*^{L530P/L530P} adult fibroblasts isolated from muscle at 4 weeks of age with death of the mutant cells by passage 13. *n* = 5 cell lines for mutant fibroblasts, 5 for heterozygotes, and 6 for wild type.

10-day course of growth. The premature death of mutant cells suggested that they either had a predetermined shortened lifespan, or that they were unable to respond adequately to the crisis stage that inevitably faces primary mouse cell lines before immortalization. In humans, ageing is associated with telomere shortening; however, we observed no difference in telomere lengths between tissues of wild-type and *Lmna*^{L530P/L530P} homozygous mice (data not shown).

Thinning of the skin, hypoplasia and degeneration of cardiac and skeletal muscle, osteoporosis, and abnormal dentition in homozygous *Lmna*^{L530P/L530P} mice are all phenotypes consistent with those seen in progeria patients¹. Marked growth retardation and shortened lifespan are also hallmarks of progeria. The question of how faithfully progeria simulates natural ageing is, however, controversial, because cataracts, age-related cancers and cognitive decline are not observed in progeria patients¹². *Lmna*^{L530P/L530P} homozygous mutant mice also failed to show any signs of these age-induced pathologies. One possibility is that progeria is a developmental disease in which there is either accelerated or improper development of certain tissue types, while other tissues are relatively unaffected. The tissues most severely affected in progeria patients and the mutant mice arise from the mesenchymal stem cell lineage, making this a particularly interesting cell lineage for further study in *Lmna*^{L530P/L530P} mice. Accelerated development and premature death of terminally differentiated mesenchymal cells may account for phenotypes in progeric heart, muscle, bone and subcutaneous adipose tissues. Alternatively, terminally differentiated tissues with defects in the nuclear lamina may be unable to maintain the chromatin organization necessary to preserve a state of terminal differentiation, resulting in dedifferentiation and subsequent redifferentiation into other cell types with age in certain tissues^{20,21}. Our preliminary data indicate that homozygous *Lmna*^{L530P/L530P} muscle myoblasts and fibroblasts readily differentiate into adipocytes, suggesting this latter hypothesis may be valid. Notably, muscles in older individuals show accumulation of fat compared with younger muscle, which could be a result of muscle tissue transdifferentiating into fat during old age²⁰. We are further studying these questions in the *Lmna*^{L530P/L530P} mouse in expectation of learning more about the extent to which differentiation and the loss of cell identity contribute to the ageing process, especially with respect to the phenotypes seen in progeria. □

Methods

Generation of *Lmna* knockin mice

Site-directed mutagenesis of a 1.4-kilobase *Bgl*II–*Bam*HI fragment was performed to introduce a nucleotide polymorphism resulting in the substitution of proline for leucine at amino acid 530 in the *Lmna* gene (see Supplementary Fig. 1A). A neomycin selectable marker flanked by *loxP* sites was inserted in reverse orientation to *Lmna* at the *Sma*I site of intron 9. The targeting vector was linearized and electroporated into W9.5 embryonic stem (ES) cells. Clones selected with neomycin were picked, expanded and screened for recombination. Two recombinant ES lines were injected into C57BL/6 blastocysts, and chimaeras were bred to produce germline offspring as described²². Homozygous, heterozygous and wild-type mice were distinguished based on subjecting an exon 9 PCR product to heteroduplex mobility analysis to discern the presence of the point mutation²³. To excise the neomycin selectable marker, *Lmna*^{L530Pneo/+} mice were crossed to mice transgenic for Cre driven by the ubiquitously expressed β 1-actin²⁴. Progeny were screened for the loss of the neomycin-resistance marker. A PCR-amplified fragment of *Lmna* exon 9 was sequenced from *Lmna*^{L530P/L530P} mice to verify the presence of the point mutation. All phenotypic characterizations were performed using *Lmna*^{L530P/L530PΔneo} mice. Most of the pathological characterization of *Lmna*^{L530P/L530P} mice was performed using a mouse line derived from one of the ES cell lines; however, no discernible differences in phenotypes of mice established from the two ES cell lines have been noted.

Characterization of mice

Pathological examination of the mice was performed at approximately 3.5–4 weeks of age. Tissues for histological analysis were fixed in 10% phosphate buffered formalin, dehydrated, cleared, embedded in paraffin, sectioned at 6 μ m, and stained with haematoxylin and eosin. Tissues sectioned included heart, pancreas, thyroid, trachea, oesophagus, liver, spleen, thymus, testis, vertebrae with spinal cord, nasal structures, tongue, periscapular fat pad, gonadal fat pad, and muscles from the hip, shoulder, femur, tibia, humerus and radius. To estimate cell number in the heart, total DNA was isolated from wild-type and mutant hearts of four mice of each genotype and normalized for the

amount of tissue extracted. Skeletal preparations were performed as described²⁵. To quantify bone mineral density, whole mice were scanned using a GE Medical Systems Lunar Pixmap densitometry scanner. Three scans per mouse were performed. Error bars indicate standard deviation.

Immunofluorescence

Indirect immunofluorescence was performed on primary MEFs at passage 2 or 3. Anti-lamin A was a gift from B. Burke, and recognizes only lamin A and not lamin C. Anti-lamin C was made to a lamin-C-specific peptide conjugated to keyhole limpet haemocyanin in rabbits. The resulting antiserum was preadsorbed with *Lmna*-null primary MEFs fixed in 3% paraformaldehyde in PBS, and the antiserum recognizes lamin C but not lamin A. Goat anti-rabbit-rhodamine secondary antibodies were used at 1:200 (Biosource International). Continuous passage assays¹⁹ were performed on adult fibroblasts isolated from muscle²⁶. For each passage, cells were plated at a density of 5×10^5 and counted every 3.5 days. Standard error of the mean is indicated in error bars.

Received 10 February; accepted 8 April 2003; doi:10.1038/nature01631.

- Uitto, J. Searching for clues to premature aging. *Trends Endocrinol. Metab.* **13**, 140–141 (2002).
- de Paula Rodriguez, G. H. et al. Severe bone changes in a case of Hutchinson-Gilford syndrome. *Ann. Genet.* **45**, 151–155 (2002).
- Burke, B. & Stewart, C. L. Life at the edge: the nuclear envelope and human disease. *Nature Rev. Mol. Cell Biol.* **3**, 575–585 (2002).
- Roeder, R.-A., Sauter, H., Weber, K. & Osborn, M. Cells of the cellular immune and hemopoietic system of the mouse lack lamins A/C: distinction versus other somatic cells. *J. Cell Sci.* **95**, 587–598 (1990).
- Hutchison, C. Lamins: building blocks or regulators of gene expression? *Nature Rev. Mol. Cell Biol.* **3**, 848–858 (2002).
- Spann, T. P., Goldman, A. E., Wang, C., Huang, S. & Goldman, R. D. Alteration of nuclear lamin organization inhibits RNA polymerase II-dependent transcription. *J. Cell Biol.* **156**, 603–608 (2002).
- Ozaki, T. et al. Complex formation between lamin A and the retinoblastoma gene product: identification of the domain on lamin A required for its interaction. *Oncogene* **9**, 2649–2653 (1994).
- Holaska, J. M., Lee, K. K., Kowalski, A. K. & Wilson, K. L. Transcriptional repressor Germ Cell-less (GCL) and Barrier to Autointegration Factor (BAF) compete for binding to Emerin *in vitro*. *J. Biol. Chem.* **278**, 6969–6975 (2003).
- Worman, H. J. & Courvalin, J. C. The nuclear lamina and inherited disease. *Trends Cell Biol.* **12**, 591–598 (2002).
- Hoeffel, J. C., Mainard, L., Chastagner, P. & Hoeffel, C. C. Mandibulo-acral dysplasia. *Skeletal Radiol.* **29**, 668–671 (2000).
- Bonne, G. et al. Mutations in the gene encoding lamin A/C cause autosomal dominant Emery-Dreifuss muscular dystrophy. *Nature Genet.* **21**, 285–288 (1999).
- Ackerman, J. & Gilbert-Barnes, E. Hutchinson-Gilford progeria syndrome: a pathologic study. *Pediatr. Pathol. Mol. Med.* **21**, 1–13 (2002).
- Sarkar, P. K. & Shinton, R. A. Hutchinson-Gilford progeria syndrome. *Postgrad. Med. J.* **77**, 312–317 (2001).
- Sullivan, T. et al. Loss of A-type lamin expression compromises nuclear envelope integrity leading to muscular dystrophy. *J. Cell Biol.* **147**, 913–920 (1999).
- Lewis, M. PRELP, collagen, and a theory of Hutchinson-Gilford progeria. *Ageing Res. Rev.* **2**, 95–105 (2003).
- Vigouroux, C. et al. Nuclear envelope disorganization in fibroblasts from lipodystrophic patients with heterozygous R482Q/W mutations in the lamin A/C gene. *J. Cell Sci.* **114**, 4459–4468 (2001).
- Morris, G. E. The role of the nuclear envelope in Emery-Dreifuss muscular dystrophy. *Trends Mol. Med.* **7**, 572–577 (2001).
- Raharjo, W. H., Enarson, P., Sullivan, T., Stewart, C. L. & Burke, B. Nuclear envelope defects associated with LMNA mutations cause dilated cardiomyopathy and Emery-Dreifuss muscular dystrophy. *J. Cell Sci.* **114**, 4447–4457 (2001).
- Todaro, G. J. & Green, H. Quantitative studies of the growth of mouse embryo cells in culture and their development into established lines. *J. Cell Biol.* **17**, 299–313 (1963).
- Taylor-Jones, J. M. et al. Activation of an adipogenic program in adult myoblasts with age. *Mech. Ageing Dev.* **123**, 649–661 (2002).
- Kirkland, J. L., Tchikonia, T., Pirtskhalava, T., Han, J. & Karagiannides, I. Adipogenesis and aging: does aging make fat go MAD? *Exp. Gerontol.* **37**, 757–767 (2002).
- Stewart, C. L. Production of chimeras between embryonic stem cells and embryos. *Methods Enzymol.* **225**, 823–855 (1993).
- Upchurch, D. A., Shankarappa, R. & Mullins, J. I. Position and degree of mismatches and the mobility of DNA heteroduplexes. *Nucleic Acids Res.* **28**, E69 (2000).
- Lewandoski, M., Meyers, E. N. & Martin, G. R. Analysis of Fgf8 gene function in vertebrate development. *Cold Spring Harb. Quant. Biol.* **62**, 159–168 (1997).
- Lufkin, T. et al. Homeotic transformation of the occipital bones of the skull by ectopic expression of a homeobox gene. *Nature* **359**, 835–841 (1992).
- Sabourin, L. A. et al. Reduced differentiation potential of primary MyoD^{-/-} myogenic cells derived from adult skeletal muscle. *J. Cell Biol.* **144**, 631–643 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank L. Sewell for technical assistance in our animal facility; D. Haines and D. Smith for help in pathological examination of the mice; H. Wimbrow for assistance in the use of the Pixmap densitometer; R. Frederickson for preparation of the figures and B. Burke and B. Howard for discussions.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.L.S. (stewartc@ncicrf.gov).

Bmi-1 is required for maintenance of adult self-renewing haematopoietic stem cells

In-kyung Park*, Dalong Qian*, Mark Kiel†, Michael W. Becker*, Michael Pihalja*, Irving L. Weissman‡, Sean J. Morrison† & Michael F. Clarke*

* Division of Hematology/Oncology, Internal Medicine, and † Howard Hughes Medical Institute, Department of Internal Medicine, and Department of Cell and Developmental Biology, University of Michigan, Ann Arbor, Michigan 48109, USA

‡ Department of Pathology, School of Medicine, Stanford University, Stanford, California 94305, USA

A central issue in stem cell biology is to understand the mechanisms that regulate the self-renewal of haematopoietic stem cells (HSCs), which are required for haematopoiesis to persist for the lifetime of the animal¹. We found that adult and fetal mouse and adult human HSCs express the proto-oncogene *Bmi-1*. The number of HSCs in the fetal liver of *Bmi-1*^{-/-} mice² was normal. In postnatal *Bmi-1*^{-/-} mice, the number of HSCs was markedly reduced. Transplanted fetal liver and bone marrow cells obtained from *Bmi-1*^{-/-} mice were able to contribute only transiently to haematopoiesis. There was no detectable self-renewal of adult HSCs, indicating a cell autonomous defect in *Bmi-1*^{-/-} mice. A gene expression analysis revealed that the expression of stem cell associated genes³, cell survival genes, transcription factors, and genes modulating proliferation including *p16*^{Ink4a} and *p19*^{Arf} was altered in bone marrow cells of the *Bmi-1*^{-/-} mice. Expression of *p16*^{Ink4a} and *p19*^{Arf} in normal HSCs resulted in proliferative arrest and p53-dependent cell death, respectively. Our results indicate that *Bmi-1* is essential for the generation of self-renewing adult HSCs.

To understand the molecular mechanisms that control the self-renewal of haematopoietic stem cells (HSCs), we have previously constructed a complementary DNA library using highly purified cells with long-term self-renewal potential⁴ and analysed the differential gene expression profile of HSCs⁵. Analysis of genes expressed in HSCs identified *Bmi-1*, a member of the *Polycomb Group* (*PcG*) family linked to malignant transformation, lymphocyte development, neurological development and senescence of fibroblasts^{2,6-8}. *Bmi-1* expression, which declines during haematopoietic development⁹, was detected at both single-cell and ten-cell levels by polymerase chain reaction with reverse transcription (RT-PCR) in mouse fetal liver and adult HSCs as well as human normal and leukaemic stem cells, indicating that HSCs express high levels of *Bmi-1* (Supplementary Fig. 1).

The *Bmi-1*^{-/-} mice develop hypocellular bone marrows and die less than 2 months after birth. To determine the function of *Bmi-1* in haematopoiesis, *Bmi-1*^{-/-} mice² in a C57Bl/Ka.Thy1.1 background¹ were analysed. As reported previously², *Bmi-1*^{-/-} mice have normal numbers of peripheral blood myeloid cells but smaller numbers of lymphocytes, whereas the frequency of colony-forming cells in the bone marrow was normal (Supplementary Fig. 2).

In vertebrates, haematopoiesis is organized as a hierarchy, in which HSCs contribute to haematopoiesis for the life of the animal and give rise to transiently reconstituting multipotent progenitors (MPPs)¹. To determine whether these mice have normal numbers of stem cells that have lost the ability to make progeny or whether they have a defect in the generation of stem cells, we investigated the frequency of HSCs and MPPs from 4–5-week-old *Bmi-1*^{+/+}, *Bmi-1*^{+/-} and *Bmi-1*^{-/-} mice by flow cytometry (Fig. 1a, b). The frequency of HSCs (Thy1.1^{low}c-kit⁺Sca-1⁺Lineage⁻ cells) in the

bone marrow of heterozygous mice was similar to that of the wild-type littermates (about 0.01% of bone marrow cells). However, the HSC frequency of the *Bmi-1*^{-/-} mice ranged from 0% to 0.013%, with an average of 0.005%. When the total numbers of bone marrow cells were taken into account, *Bmi-1*^{-/-} mice had an average 10-fold (up to 45-fold) less total HSCs as measured by flow cytometry (*P* = 0.007; Supplementary Fig. 3). The frequency of MPP cells (Thy1.1^{low}c-kit⁺Sca-1⁺Lineage^{-/low} cells, which contain both HSCs and a majority population of MPPs) in the *Bmi-1*^{-/-} bone marrow was similar to that of wild-type mice (about 0.1% of bone marrow cells). Next, competitive repopulation experiments¹ were performed to determine the long-term reconstituting abilities of bone marrow mononuclear cells obtained from 5-week-old *Bmi-1*^{-/-} mice. At 5 weeks after transplantation, flow cytometry analysis of peripheral blood revealed that bone marrow cells from both wild-type and heterozygous mice efficiently reconstituted myeloid cells, B cells and T cells, whereas *Bmi-1*^{-/-} bone marrow cells showed a decreased capacity to reconstitute all lineages in the transplanted animals (Fig. 1c). By 10 weeks, the mice reconstituted with *Bmi-1*^{-/-} marrow, but not with wild-type bone marrow, lost nearly all donor-derived mature haematopoietic cells (Fig. 1d). These results indicate that *Bmi-1* might be required for self-renewal and/or maintenance of adult HSCs.

The defect in HSCs in the bone marrow of *Bmi-1*^{-/-} mice could result from an impaired generation of self-renewing fetal and/or adult HSCs, defective stem cell homing, or a defective bone marrow microenvironment. We therefore next examined the HSC compartment in fetal *Bmi-1*^{-/-} mice. As measured by flow cytometry, *Bmi-1*^{-/-} mice had similar numbers of fetal liver cells and an almost identical frequency of fetal liver HSCs¹⁰ (Thy1.1^{low}Mac-1⁺Sca-1⁺Lineage⁻) as their wild-type littermates (Fig. 2a). Furthermore, the fetal liver HSCs from *Bmi-1*^{-/-} mice migrated normally towards the chemokine SDF-1 α , and fetal liver cells expressed normal levels of SDF-1 α and its receptor CXCR4, important mediators of HSC homing¹¹ (Supplementary Fig. 4).

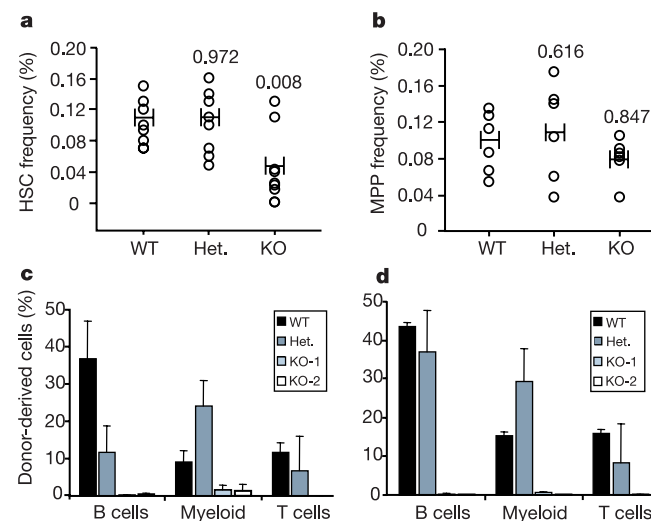


Figure 1 Analysis of adult HSCs. **a, b**, Effect of *Bmi-1* deletion on the frequencies of HSCs and MPPs. The bars show the average frequencies of HSCs (**a**) and MPPs (**b**) from eight individual mice in each group. *P* values (shown above the bars) were calculated with the unpaired Student's *t*-test. WT, wild type; Het., *Bmi-1*^{+/-}; KO, *Bmi-1*^{-/-}. **c, d**, Competitive reconstitution. Donor (Ly5.1) bone marrow cells (5×10^5) were mixed with the same number of Ly5.2 bone marrow cells and injected into lethally irradiated Ly5.2 mice (*n* = 2–4). Peripheral blood was analysed 5 weeks (**c**) and 10 weeks (**d**) after reconstitution for donor-derived myeloid, B-lymphoid and T-lymphoid cells. KO-1 and KO-2, *Bmi-1*^{-/-} mice 1 and 2.

Although it is possible that *Bmi-1* affects homing by other pathways, these data suggest that the HSC defect in *Bmi-1*^{-/-} bone marrow is not likely to be the result of defective fetal liver HSC migration to the bone marrow.

If the defects in the *Bmi-1*^{-/-} mice were secondary to an abnormal microenvironment, then fetal liver stem cells from these mice would be able to engraft the bone marrow of wild-type mice efficiently. If the defects were in the generation of self-renewing HSCs, then *Bmi-1*^{-/-} fetal liver HSCs would fail to permanently engraft the bone marrow of normal adult mice. To distinguish between these possibilities, lethally irradiated Ly5.2 congenic mice were transplanted with 5×10^5 , 10^6 and 5×10^6 *Bmi-1*^{-/-} cells, or 5×10^5 and 10^6 *Bmi-1*^{+/+} embryonic day 14.5 (E14.5) fetal liver cells. In addition, competitive reconstitution assays were performed with 10^6 fetal liver cells from *Bmi-1*^{+/+} or *Bmi-1*^{-/-} mice mixed with 10^6 Ly5.2 bone marrow cells. At 4 weeks after transplantation, *Bmi-1*^{-/-} mice showed reconstitution of T cells and myeloid cells, whereas B-cell reconstitution was greatly decreased (Fig. 2b). By 8 weeks, virtually no donor-derived cells were detected in recipient bone marrows reconstituted with *Bmi-1*^{-/-} fetal liver cells (Fig. 2c), suggesting a severe defect in the HSCs' ability to maintain haematopoiesis. At 8 weeks after the original transplant, secondary bone marrow transplants were performed with 10^7 donor cells. Peripheral blood analysis after 6 weeks indicated that there were no detectable *Bmi-1*^{-/-} cells in the mice with secondary transplants (Fig. 2d). Taken together with

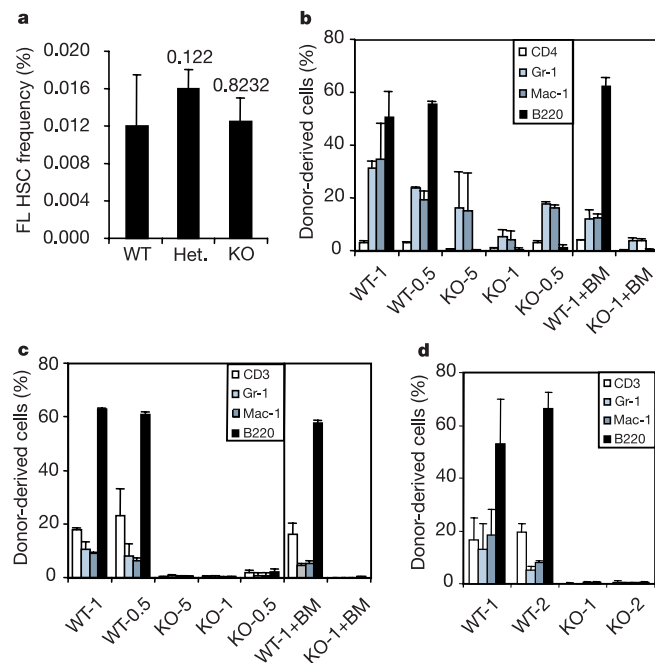


Figure 2 Analysis of fetal liver haematopoietic stem cells in *Bmi-1*^{+/+}, *Bmi-1*^{+/-} and *Bmi-1*^{-/-} mice. **a**, The fetal liver HSC population was analysed from two *Bmi-1*^{+/+} (WT), three *Bmi-1*^{+/-} (Het.) and four *Bmi-1*^{-/-} (KO) fetuses by flow cytometry. *P* values (shown above the bars) were calculated with the unpaired Student's *t*-test.

b, c, Peripheral blood analysis of lethally irradiated mice reconstituted with 10^6 and 5×10^5 *Bmi-1*^{+/+} fetal liver cells (WT-1 and WT-0.5, respectively) and 5×10^6 , 10^6 and 5×10^5 *Bmi-1*^{-/-} fetal liver cells (KO-5, KO-1 and KO-0.5, respectively). WT-1 + BM and KO-1 + BM indicate competitive reconstitution with 10^6 fetal liver cells together with 10^6 Ly5.2 bone marrow cells (five mice in each group). Data for 4 weeks (**b**) and 8 weeks (**c**) after reconstitution are shown. **d**, Secondary bone marrow transfer. After 8 weeks of reconstitution, 10^7 bone marrow cells from mice non-competitively reconstituted with 10^6 fetal liver cells were used to transplant lethally irradiated Ly5.2 mice. Peripheral blood analysis was performed 6 weeks after transplant.

the marked decrease in HSC numbers in the bone marrow of the *Bmi-1*^{-/-} mice, these results demonstrate that there is a profound intrinsic defect in the maintenance of self-renewing HSCs in the mutant mice.

To begin to understand the molecular mechanism by which *Bmi-1* affects the generation of HSCs, we compared the gene expression profiles of bone marrow mononuclear cells obtained from wild-type and *Bmi-1*^{-/-} mice. *Bmi-1* is a member of the *PcG* family of transcriptional repressors that control development by the regulation of cell growth and differentiation genes¹². Microarray analysis with bone marrow RNAs isolated from wild-type and *Bmi-1*^{-/-} mice identified several interesting genes (Fig. 3a, b and Supplementary Table 2). These include wild-type p53-induced gene 1 (*Wig1*), tight-junction protein 1 (TJP1) and platelet-activating factor acetylhydrolase isoform 1b (PAF-AHγ), all of which are expressed in embryonic, neuronal and haematopoietic stem cells³, as well as several transcription factors, signalling proteins and apoptosis inhibitor 6. As reported previously in embryonic fibroblasts, expression of *p16*^{Ink4a} and *p19*^{Arf}, which are generated by alternative splicing from the *Ink4a* locus^{13–15}, in adult bone marrow of *Bmi-1*^{-/-} mice was elevated compared with wild-type littermates (Fig. 3a). A complete list of differentially expressed genes is given in Supplementary Table 3. In contrast with the analysis of the adult bone marrow, a microarray analysis of E14.5 fetal liver cells did not show significant differences in the gene expression by the wild-type and mutant cells (data not shown), which is consistent with the observation that *Bmi-1* is not crucial for fetal liver haematopoiesis. Some members of the *Hox* family are also thought to be targets of *Bmi-1*. Semiquantitative RT-PCR indicated that expression of *Hoxb4*, which is important for HSC development¹⁶, was not altered in *Bmi-1*^{-/-} mice whereas *Hoxa9*, which functions in cell fate decisions of haematopoietic progenitors¹⁷, was slightly upregulated in thymus and bone marrow of *Bmi-1*^{-/-} mice (Supplementary Fig. 5).

Next, the effects of the *Bmi-1* targets, *p16*^{Ink4a} and *p19*^{Arf}, on HSC function were determined. HSCs (c-Kit⁺Sca-1⁺Flt3⁻(Flk-2) Lineage⁻ phenotype¹⁸) were isolated and infected with mouse stem cell virus (MSCV) expressing green fluorescent protein (GFP) alone or together with *p16*^{Ink4a} or *p19*^{Arf}. Two days after infection, only 6% of cells infected with the control virus and the *p16*^{Ink4a} virus, compared with 60% of the cells infected with the *p19*^{Arf} virus, had lost viability as measured by exclusion of a viability dye 7AAD (Fig. 4). When single viable GFP-positive HSCs were placed in culture medium, 79% of the haematopoietic progenitors infected with the control GFP virus generated more than 50 cells after 7 days of culture (75 of 95 wells seeded). By contrast, progenitors infected with the *p16*^{Ink4a} virus either did not prolifer-

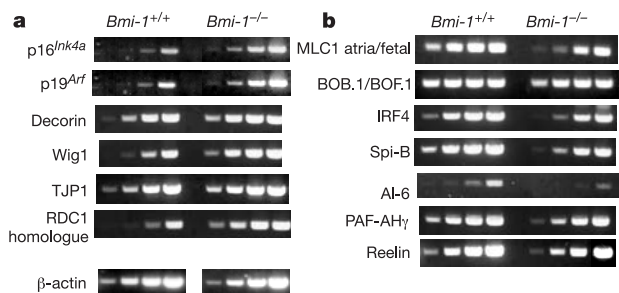


Figure 3 Regulation of gene expression by *Bmi-1*. Semiquantitative RT-PCR of genes upregulated (**a**) and downregulated (**b**) in *Bmi-1*^{-/-} (KO) mice. cDNA was reverse transcribed from 2 μg of total RNA, and gene-specific primers were used for PCR. Aliquots were taken at the end of the 26th, 29th, 32nd and 35th cycles. The β-actin aliquots were taken at the end of the 17th, 20th, 23rd and 26th cycles.

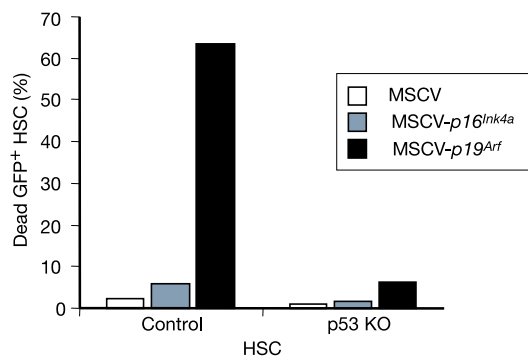


Figure 4 Effects of *p16^{Ink4a}* and *p19^{Arf}* on HSC proliferation. HSCs from wild-type and *p53*^{-/-} (KO) mice were infected with various MSCV viruses. After 48 h, cells were analysed for GFP expression by flow cytometry. Percentages of GFP-positive 7AAD-positive (dead transduced) cells are shown.

ate at all or generated only a few cells (3–10 cells, in 3 of 95 wells). None of the *p19^{Arf}* virus-infected HSCs proliferated (0 of 47 wells). The *p19^{Arf}* protein increases the level of p53 protein¹⁹. To determine whether *p19^{Arf}*-induced cell death is mediated by p53, we infected *p53*^{-/-} HSCs (c-Kit⁺Sca-1⁺Flt3⁻Lineage⁻ phenotype) with the *p19^{Arf}* virus. In contrast with the wild-type HSCs, these cells were viable 2 days after infection (Fig. 4). These results suggest that *Bmi-1* targets *p16^{Ink4a}* to allow the proliferation of HSCs and *p19^{Arf}* to inhibit apoptosis through p53.

HSCs undergo self-renewing cell divisions for the lifetime of an animal. Once an HSC has undergone cell division, the daughter cells must choose one of three possible fates: to remain as a stem cell, to differentiate into progenitors or to undergo apoptosis (Fig. 5). Although several genes have been shown to be important for stem cell function, the mechanisms that regulate the self-renewal of HSCs are poorly understood^{20–26}. Our results suggest that *Bmi-1* is necessary for efficient self-renewing cell divisions of adult HSCs but is less critical for the generation of differentiated progeny (Fig. 5). Transplantation of *Bmi-1*^{-/-} fetal liver cells resulted in transient haematopoietic cell reconstitution. This indicates that the transplanted mutant fetal liver HSCs might not have generated HSCs efficiently but might have given rise to MPPs that could sustain haematopoiesis for only 4–8 weeks. This could have been due to the expression of other *PcG* genes by the MPPs. Consistent with this model was our previous result that MPPs express more *Enx-1* than do HSCs⁵. Similarly, other *PcG* genes, such as *rae28*, whose expression is necessary for fetal liver haematopoiesis to occur²⁷, might permit HSC proliferation in the fetus of the mutant mice. This is consistent with the observation that the expression of *p19^{Arf}* is elevated in adult bone marrow cells, but not in fetal liver cells, of the mutant mice (Supplementary Fig. 5). Alternatively, the number of self-renewing cell division of *Bmi-1*^{-/-} FL-HSCs might be restricted.

The mechanism by which *Bmi-1* modulates HSC self-renewal seems to be through the regulation of genes important for stem cell fate decisions as well as that of survival genes, anti-proliferative genes and stem-cell-associated genes. The gene expression data and retroviral infection experiments implicate *p16^{Ink4a}*, *p19^{Arf}* and p53 as downstream effectors of *Bmi-1* involved in the control of the proliferation and survival of HSCs during self-renewing cell divisions (Fig. 5). However, owing to the paucity of HSCs in the *Bmi-1*^{-/-} mice, it is not possible to determine whether the expression levels of these genes are comparable to those of HSCs infected with the *p16^{Ink4a}* and *p19^{Arf}* retroviruses. Although the increased expression of the p53-target gene *Wig1* suggests that *p19^{Arf}* is activated in *Bmi-1*^{-/-} haematopoietic cells, the relative contribution of each of these pathways to the regulation of self-renewing HSC generation awaits careful analysis of the HSCs of

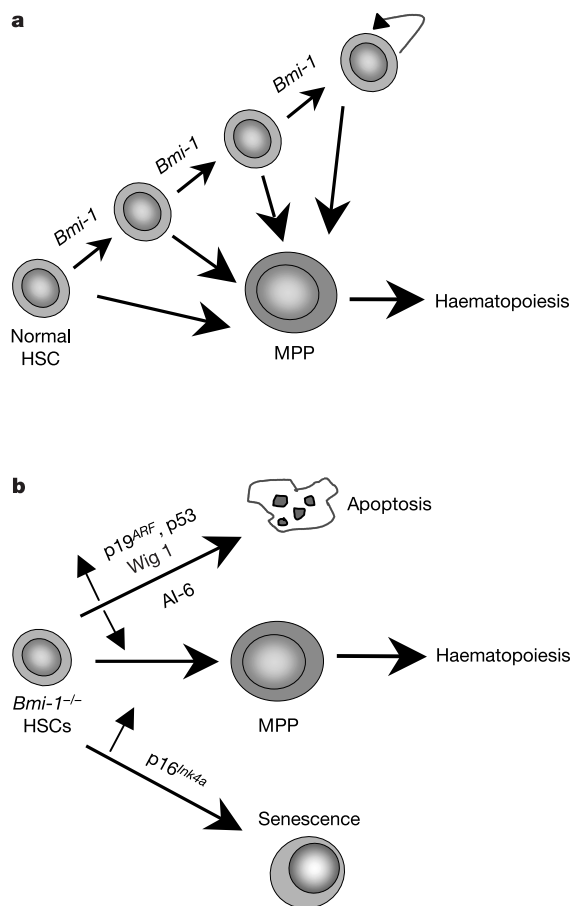


Figure 5 Proposed model for the role of *Bmi-1* in HSC maintenance. **a**, In wild-type mice, a HSC can undergo asymmetric cell division to generate another HSC as well as a MPP, thereby maintaining the HSC pool. **b**, In the absence of *Bmi-1*, the increased expression of *p16^{Ink4a}* can induce a proliferative arrest and increased expression of *p19^{Arf}* can increase the level of p53 and its downstream target, *Wig1*, which, coupled with a decreased expression of apoptosis inhibitor 6 (*Al-6*), results in the death of HSCs. This results in severely impaired adult HSC self-renewal. By contrast, MPPs, which are initially generated from the HSCs, transiently give rise to mature haematopoietic cells.

Bmi-1^{-/-} mice that also harbour mutations of one or more of these other genes. Some of the other genes identified in the microarray analysis might also have crucial roles in stem cell biology. For example, *Wig1* and apoptosis inhibitor 6 might regulate the survival of HSCs. As *Bmi-1* has now been shown³⁰ to be necessary for the maintenance of mouse leukaemia and *Bmi-1* is expressed by human AML stem cells, it will be of interest to determine whether *Bmi-1* has a similar function in human leukaemia. □

Methods

Flow cytometry

Bone marrow cells were obtained by flushing the tibias and femurs of mice, and stained with antibodies against lineage markers (CD3, CD4, CD5, CD8, Gr-1, Mac-1 and Ter119), Sca-1, c-Kit and Thy1.1 as described¹. Cells were analysed by flow cytometry with a Vantage fluorescence-activated cell sorter (Becton Dickinson). The phenotype of adult HSCs, multipotent progenitors (MPPs) and fetal liver HSCs are Thy1.1^{low}c-Kit⁺Sca-1⁺Lineage⁻, Thy1.1^{low}c-Kit⁺Sca-1⁺Lineage⁻ and Thy1.1^{low}Sca-1⁺Mac-1⁺Lineage⁻, respectively. When Thy1.2 background mice (C57BL/6J p53-deficient mice and C57BL/6J mice) were used, HSCs and MPPs were isolated as c-Kit⁺Sca-1⁺Flt3⁻Lineage⁻ and c-Kit⁺Sca-1⁺Flt3⁺Lineage⁻ cells, respectively¹⁸.

Competitive reconstitution

Lethally irradiated C57BL/Ka-Ly5.2 congenic mice were competitively reconstituted with whole bone marrow or E14.5 fetal liver cells from *Bmi-1*^{+/+}, *Bmi-1*^{+/-} and *Bmi-1*^{-/-}

mice with or without the C57Bl/Ka-Ly5.2 recipient bone marrow cells¹. Reconstitution of donor (Ly5.1) myeloid and lymphoid cells was monitored by staining blood cells with antibodies against Ly5.1, CD3, B220, Mac-1 and Gr-1. The secondary bone marrow transplant was performed with 10⁷ whole bone marrow cells from mice reconstituted with *Bmi-1*^{+/-} or *Bmi-1*^{-/-} fetal liver cells.

Retroviral gene transfer of HSCs

Mouse stem cell viruses expressing mouse *p16*^{Ink4a} or *p19*^{Arf} cDNAs together with GFP were produced using Phoenix ecotropic packaging cells²⁸. Infection of HSCs was done as described²⁹ except that three cycles of infections were performed. After 48 h, single GFP-positive cells were sorted into a 96-well plate containing 100 µl HSC medium²⁹ and grown for 7 days. Each well was scored for the presence of GFP-positive cells by observation with a fluorescence microscope.

Received 10 February; accepted 19 March 2003; doi:10.1038/nature01587.

Published online 20 April 2003.

- Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661–673 (1994).
- van der Lugt, N. M. *et al.* Posterior transformation, neurological abnormalities, and severe hematopoietic defects in mice with a targeted deletion of the *bmi-1* proto-oncogene. *Genes Dev.* **8**, 757–769 (1994).
- Ramalho-Santos, M. *et al.* 'Stemness': transcriptional profiling of embryonic and adult stem cells. *Science* **298**, 597–600 (2002).
- Park, I.-K. *et al.* Molecular cloning and characterization of a novel regulator of G-protein signaling from mouse hematopoietic stem cells. *J. Biol. Chem.* **276**, 915–923 (2001).
- Park, I. K. *et al.* Differential gene expression profiling of adult murine hematopoietic stem cells. *Blood* **99**, 488–498 (2002).
- Lessard, J., Baban, S. & Sauvageau, G. Stage-specific expression of Polycomb group genes in human bone marrow cells. *Blood* **91**, 1216–1224 (1999).
- Kiyono, T. *et al.* Both Rb/p16^{INK4a} inactivation and telomerase activity are required to immortalize human epithelial cells. *Nature* **396**, 84–88 (1998).
- van der Lugt, N. M. T., Alkema, M., Berns, A. & Deschamps, J. The Polycomb-group homolog *Bmi-1* is a regulator of murine Hox gene expression. *Mech. Dev.* **58**, 153–164 (1996).
- Akashi, K. *et al.* Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383–389 (2003).
- Morrison, S., Hemmati, H., Wandycz, A. & Weissman, I. The purification and characterization of fetal liver hematopoietic stem cells. *Proc. Natl Acad. Sci. USA* **92**, 10302–10306 (1995).
- Wright, D. E. *et al.* Hematopoietic stem cells are uniquely selective in their migratory response to chemokines. *J. Exp. Med.* **195**, 1145–1154 (2002).
- Mahmoudi, T. & Verrijzer, C. P. Chromatin silencing and activation by Polycomb and trithorax group proteins. *Oncogene* **20**, 3055–3066 (2001).
- Weber, J. D. *et al.* Nucleolar Arf sequesters Mdm2 and activates p53. *Nature Cell Biol.* **1**, 20–26 (1999).
- Jacob, J. *et al.* The oncogene and Polycomb-group gene *bmi-1* regulates cell proliferation and senescence through the *ink4a* locus. *Nature* **397**, 164–168 (1999).
- Quelle, D. E., Zindy, F., Ashmun, R. A. & Sherr, C. J. Alternative reading frames of the *INK4a* tumour suppressor gene encode two unrelated proteins capable of inducing cell cycle arrest. *Cell* **84**, 993–1000 (1995).
- Antonchuk, J., Sauvageau, G. & Humphries, R. K. HOXB4 overexpression mediates very rapid stem cell regeneration and competitive hematopoietic repopulation. *Exp. Hematol.* **29**, 1125–1134 (2002).
- Lawrence, H. J. *et al.* Mice bearing a targeted interruption of the homeobox gene *HOXA9* have defects in myeloid, erythroid, and lymphoid hematopoiesis. *Blood* **89**, 1922–1930 (1997).
- Christensen, J. L. & Weissman, I. L. Flk-2 is a marker in hematopoietic stem cell differentiation: a simple method to isolate long-term stem cells. *Proc. Natl Acad. Sci. USA* **98**, 14541–14546 (2001).
- Zhang, Y., Xiong, Y. & Yarbrough, W. G. ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumour suppression pathways. *Cell* **92**, 725–734 (1998).
- Shivdasani, R., Mayer, E. & Orkin, S. Absence of blood formation in mice lacking the T-cell leukaemia oncogene *tal-1*/SCL. *Nature* **373**, 432–434 (1995).
- Porcher, C. *et al.* The T cell leukemia oncogene *SCL*/tal-1 is essential for development of all hematopoietic lineages. *Cell* **86**, 47–57 (1996).
- Antonchuk, J., Sauvageau, G. & Humphries, R. K. HOXB4-induced expansion of adult hematopoietic stem cells *ex vivo*. *Cell* **109**, 39–45 (2002).
- Domen, J., Cheshier, S. H. & Weissman, I. L. The role of apoptosis in the regulation of hematopoietic stem cells: overexpression of BCL-2 increases both their number and repopulation potential. *J. Exp. Med.* **191**, 253–264 (2000).
- Nichogiannopoulou, A. *et al.* Defects in hematopoietic stem cell activity in Ikaros mutant mice. *J. Exp. Med.* **190**, 1201–1214 (1999).
- Fisher, R. C., Lovelock, J. D. & Scott, E. W. A critical role for PU.1 in homing and long-term engraftment by hematopoietic stem cells in the bone marrow. *Blood* **94**, 1283–1290 (1999).
- Cheng, T. *et al.* Hematopoietic stem cell quiescence maintained by p21^{cip1/waf1}. *Science* **287**, 1804–1808 (2000).
- Ohta, H. *et al.* Polycomb group gene *rae28* is required for sustaining activity of hematopoietic stem cells. *J. Exp. Med.* **195**, 759–770 (2002).
- Pear, W., Nolan, G., Scott, M. & Baltimore, D. Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl Acad. Sci. U.S.A.* **90**, 8392–8396 (1993).
- Cotta, C., Swindle, C., Weissman, I. L. & Klug, C. A. Retroviral Transduction of FACS-Purified Hematopoietic Stem Cells (eds Klug, C. A. & Jordan, C. T.) 243–252 (Humana Press, Totowa, New Jersey, 2001).
- Lessard, J. & Sauvageau, G. *Bmi-1* determines the proliferative capacity of normal and leukaemic stem cells. *Nature* advance online publication, 20 April 2003 (doi: 10.1038/nature01572).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank T. Magnuson and C. Klug for providing *Bmi-1*^{+/-} mice and the MSCV plasmid, respectively; and the Flow Cytometry Core and the Microarray Core at the University of Michigan for their work. The Microarray Core is supported in part by a University of Michigan's Cancer Center Support Grant from the NIH. This work is supported by grants from the NIH.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.F.C. (mclarke@umich.edu).

DNA helicase Srs2 disrupts the Rad51 presynaptic filament

Lumir Krejci^{*†}, Stephen Van Komen^{*†}, Ying Li[‡], Jana Villemain^{*}, Mothe Sreedhar Reddy^{*}, Hannah Klein[§], Thomas Ellenberger[‡] & Patrick Sung^{*}

^{*} Institute of Biotechnology and Department of Molecular Medicine, University of Texas Health Science Center at San Antonio, 15355 Lambda Drive, San Antonio, Texas 78245, USA

[‡] Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA

[§] Department of Biochemistry, New York University School of Medicine, New York, New York 10016, USA

[†] These authors contributed equally to the work

Mutations in the *Saccharomyces cerevisiae* gene *SRS2* result in the yeast's sensitivity to genotoxic agents, failure to recover or adapt from DNA damage checkpoint-mediated cell cycle arrest, slow growth, chromosome loss, and hyper-recombination^{1,2}. Furthermore, double mutant strains, with mutations in DNA helicase genes *SRS2* and *SGS1*, show low viability that can be overcome by inactivating recombination, implying that untimely recombination is the cause of growth impairment^{1,3,4}. Here we clarify the role of *SRS2* in recombination modulation by purifying its encoded product and examining its interactions with the Rad51 recombinase. *Srs2* has a robust ATPase activity that is dependent on single-stranded DNA (ssDNA) and binds Rad51, but the addition of a catalytic quantity of *Srs2* to Rad51-mediated recombination reactions causes severe inhibition of these reactions. We show that *Srs2* acts by dislodging Rad51 from ssDNA. Thus, the attenuation of recombination efficiency by *Srs2* stems primarily from its ability to dismantle the Rad51 presynaptic filament efficiently. Our findings have implications for the basis of Bloom's and Werner's syndromes, which are caused by mutations in DNA helicases and are characterized by increased frequencies of recombination and a predisposition to cancers and accelerated ageing⁵.

We have been unable to overexpress *Srs2* protein significantly in yeast, suggesting that this protein is unstable in, and/or toxic to, yeast cells. We therefore turned to *Escherichia coli* and an inducible T7 promoter as vehicle for *Srs2* expression. *Srs2* could be revealed by Coomassie Blue staining of *E. coli* extracts and by immunoblotting with antibodies against *Srs2* (Fig. 1a). We subjected *E. coli* lysate to precipitation with ammonium sulphate and a five-step chromatographic fractionation scheme to purify *Srs2* to near-homogeneity (Fig. 1b). Purified *Srs2* has a robust ssDNA-dependent ATPase activity ($k_{cat} \geq 2,500 \text{ min}^{-1}$) and a DNA helicase activity⁶ that is fuelled by ATP hydrolysis (Fig. 1c).

Previous studies have unveiled an anti-recombination function in *SRS2* and a genetic interaction with *RAD51* (refs 7–9). We investigated whether *Srs2* protein interacts physically with Rad51 protein, and also tested its effect on the Rad51 recombinase activity¹⁰. To

examine whether Srs2 associates with Rad51, we coupled the latter to Affi-gel 15 beads and used the resulting matrix to bind Srs2. As shown in Fig. 1d, Srs2 was retained on the Affi-Rad51 beads, but no binding of Srs2 to bovine serum albumin (BSA) immobilized on Affi-beads (Affi-BSA) was detected. An interaction between Rad51 and two carboxy-terminal fragments of Srs2 was seen in the two-hybrid assay in yeast (Fig. 1e). We were unable to detect significant interaction between Rad51 and full-length Srs2 in this assay, which could be the result of low expression of full-length Srs2.

We next tested the effect of Srs2 on the Rad51-mediated homologous pairing and strand exchange reaction that serves to join recombining DNA molecules^{10,11}. For this, we employed a commonly used assay in which Rad51 and the heterotrimeric ssDNA-binding factor RPA are incubated with ssDNA and ATP to form a Rad51-ssDNA nucleoprotein filament¹¹⁻¹³. Such a filament, often called the presynaptic filament¹¹⁻¹³, is then incubated with the homologous linear duplex (Fig. 2Aa). Pairing between the DNA substrates yields a joint molecule, which is further processed by DNA strand exchange to nicked circular duplex (Fig. 2Aa, b). As shown in Fig. 2Ac, d, the addition of a catalytic quantity of Srs2 strongly suppressed the homologous pairing and strand exchange reaction.

To characterize its anti-recombination activity further, Srs2 was added to D-loop reactions in which pairing of a ³²P-labelled 90-mer oligonucleotide with a homologous duplex target is mediated by the combination of Rad51 and Rad54 proteins^{14,15} (Fig. 2Ba). As reported previously¹⁴⁻¹⁶, efficient D-loop formation was catalysed by Rad51 and Rad54 (Fig. 2Bb, d). As expected, the inclusion of Srs2 decreased the level of D-loop formation (Fig. 2Bb, d). RPA enhanced D-loop formation in the absence of Srs2 (Fig. 2Bc, d), but the inhibitory effect of Srs2 became much more pronounced when RPA was present. We provide an explanation below for this observation.

We considered the possibility that suppression of recombination by Srs2 might result from the dissociation of DNA joints by its helicase activity. To address this, D-loop was formed with Rad51/Rad54 and then Srs2 was added. Srs2 was incapable of dissociating the preformed D-loop, regardless of the presence or absence of RPA

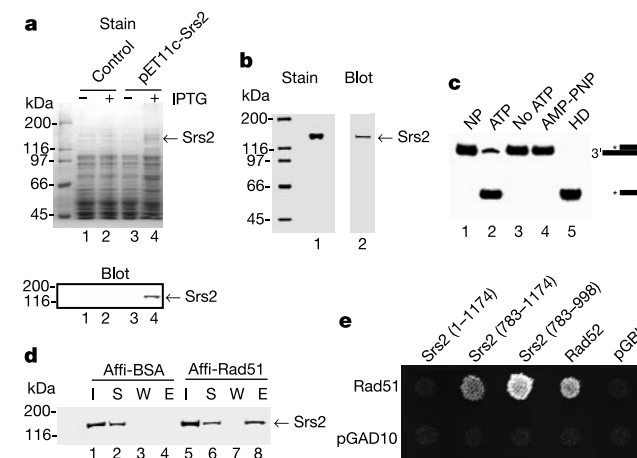


Figure 1 Purification and characterization of Srs2. **a**, Extracts from *E. coli* cells harbouring pET11c::Srs2 and the control vector pET11c grown with or without isopropyl β -D-thiogalactoside (IPTG) were analysed by SDS-PAGE and immunoblotting. **b**, Purified Srs2 was analysed by SDS-PAGE (2 μ g) and immunoblotting (20 ng). **c**, DNA unwinding by Srs2 occurs with ATP but not without it or with AMP-PNP. The substrate was also incubated alone (NP) or boiled (HD) for 1 min. **d**, Srs2 was mixed with Affi-Rad51 and Affi-BSA beads. The input (I), supernatant (S), wash (W) and SDS eluate (E) were immunoblotted. **e**, Full-length and truncated versions of Srs2 were tested for two-hybrid interaction with Rad51. Empty vectors and Rad52 were included as controls.

(Fig. 3A), suggesting that suppression of the recombination reaction does not stem from the unwinding of DNA by Srs2.

Several approaches were used to test the idea that Srs2 inhibits Rad51 recombinase function by disrupting the presynaptic filament. In doing so, we reasoned that disruption of the presynaptic filament would yield free Rad51 molecules that could be trapped on duplex DNA (dsDNA). The binding of Rad51 to topologically relaxed dsDNA induces lengthening of the DNA^{17,18} that can be monitored as a change in the DNA linking number on treatment with topoisomerase I (Fig. 3Ba). The product of this reaction is an underwound species referred to as form U (Fig. 3Bb, lane 4). RPA and Srs2 do not catalyse the formation of form U (Fig. 3Bb, lane 8), and these proteins have no effect on the formation of form U by Rad51 (Fig. 3Bb, compare lanes 6 and 4). The presynaptic filament consisting of Rad51-ssDNA does not make form U (Fig. 3Bc, lane 3). The addition of Srs2 to the Rad51-ssDNA presynaptic filament causes the generation of form U (Fig. 3Bc, lanes 4-6), indicating the transfer of Rad51 from the presynaptic filament to the dsDNA. The addition of RPA further stimulates the Srs2-mediated release of Rad51 from the presynaptic filament and the formation of form U (Fig. 3Bc, lanes 8-10). RPA has high affinity for ssDNA and it can compete with Rad51 for binding to ssDNA^{10,19,20}. The enhanced production of form U was therefore probably due to the sequestering of ssDNA by RPA after Srs2 had released Rad51, thereby preventing the renucleation of Rad51 on the ssDNA. This premise was verified by electron microscopy and explains the

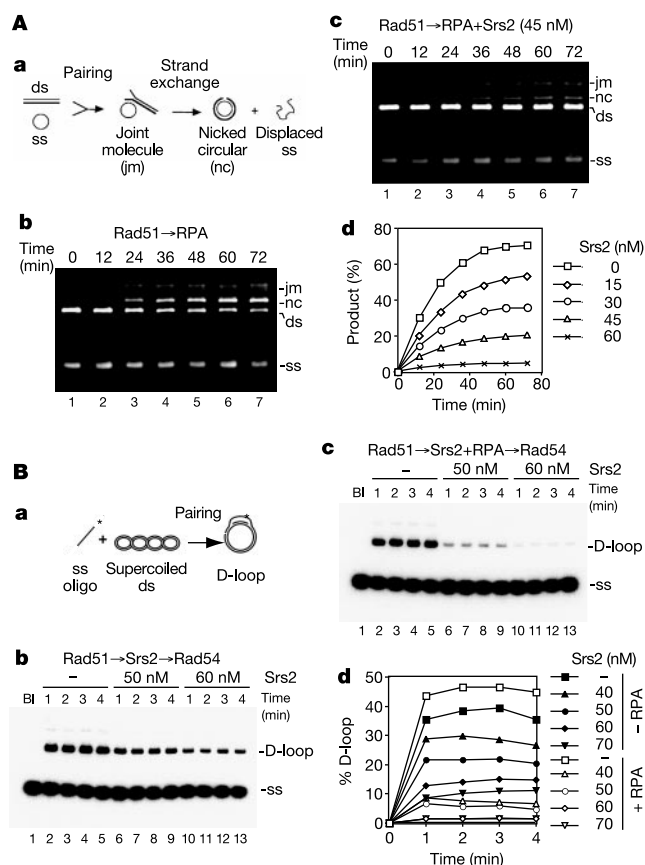


Figure 2 Srs2 inhibits Rad51-mediated DNA pairing and strand exchange. **a**, The DNA strand exchange scheme. In **b**, the DNA substrates were incubated with Rad51 and RPA. In **c**, Srs2 was also included. The results from **b** and **c** and from reactions with other Srs2 amounts are plotted in **d**. **B**, **a**, The D-loop reaction scheme. In **b**, Rad51, Srs2 and Rad54 were incubated with the DNA substrates. In **c**, RPA was also included. The results from **b** and **c** and from reactions with other Srs2 amounts are plotted in **d**. Filled symbols, reactions without RPA; open symbols, reactions with RPA.

The Srs2-mediated disruption of the Rad51 presynaptic filament was examined by a second approach. Here, Rad51 that had been dissociated from ssDNA by Srs2 was trapped on a DNA duplex bound to magnetic beads through a biotin–streptavidin linkage (Fig. 3Ca). Rad51 was eluted from the bead-bound DNA duplex by treatment with SDS and then analysed in a denaturing polyacrylamide gel. Consistent with results from the topoisomerase I-linked assay (Fig. 3B) was the observation that there was an Srs2-concentration-dependent transfer of Rad51 from the presynaptic filament to the bead-bound DNA duplex (Fig. 3Cb).

Even though homologous recombination is important for repairing DNA strand breaks induced by ionizing radiation and endogenous agents, and for restarting delinquent DNA replication forks, it can also generate deleterious genomic rearrangements and create DNA structures that cannot be properly resolved¹. Cells have therefore evolved mechanisms to avoid untimely recombination^{1,5}. Our studies provide evidence that Srs2 does this by disrupting the Rad51 presynaptic filament. The same conclusion has been reached independently²¹. Furthermore, even though RPA can function as a cofactor in the assembly of the Rad51 presynaptic filament^{10,20}, it might also promote the anti-recombination function of Srs2 by preventing reassembly of the presynaptic filament (Figs 3 and 4). That Srs2 uses the free energy from ATP hydrolysis to dislodge Rad51 from the presynaptic filament has been verified with mutant variants of Srs2 (srs2 K41A and srs2 K41R) defective for ATP hydrolysis. The physical interaction noted between Rad51 and Srs2 further suggests a mechanism for targeting the latter to the presynaptic filament. Taken together, the results presented here indicate that the motor activity of Srs2 driven by ATP hydrolysis is capable of dissociating not only DNA structures⁶ but also DNA-protein complexes.

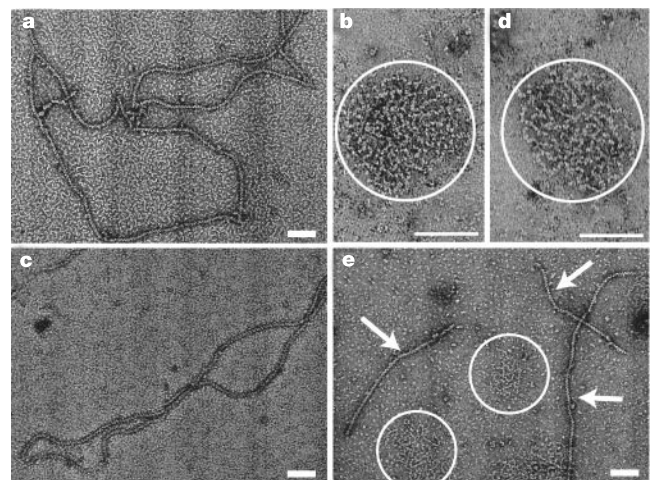


Figure 4 EM analysis of Rad51 filament disruption by Srs2. **a, b**, Rad51 (**a**) and RPA (**b**) were each incubated with ssDNA; examples of the nucleoprotein complexes that formed are shown. **c**, RPA was not able to disrupt preformed Rad51 filaments; an example of the Rad51 filaments present is shown. **d**, Incubation of preformed Rad51 filaments with Srs2 and RPA caused the loss of filaments and concomitant formation of RPA–ssDNA complexes, an example of which is shown. **e**, When preformed Rad51–ssDNA filaments were incubated with Srs2, RPA and linear duplex, RPA–ssDNA complexes were formed (circled) and transfer of Rad51 onto the linear duplex was visualized (arrows). Scale bars, 100 nm.

The inhibitory effect of Srs2 on Rad51-mediated DNA strand exchange can be partly overcome by the inclusion of Rad52 protein, a recombination mediator that promotes Rad51 presynaptic filament assembly^{11,12}. However, the Rad55–Rad57 complex, which also has recombination mediator activity^{11,12}, is much less effective in alleviating the inhibitory effect of Srs2. Furthermore, Rad52 and the Rad55–Rad57 complex do not seem to act synergistically. Interestingly, we have found that Srs2 can also dismantle presynaptic filaments of RecA and human Rad51 proteins. It therefore seems that the presynaptic filaments formed by the RecA/Rad51 class of general recombinases share a conserved feature that is recognized by Srs2, making them prone to disruption by the motor activity of Srs2.

The sensitivity of *srs2* mutants to DNA-damaging agents^{6,22} is alleviated by deleting *RAD51* (ref. 23), and the inability to remove Rad51 from DNA in the *srs2* mutants most probably accounts for the hyper-recombination phenotype of these mutants¹. Similarly, the cell cycle checkpoint recovery and adaptation defect in *srs2* mutants might be related to an inability to evict Rad51 from damaged DNA². Cells mutated for *SRS2* grow slowly, exhibit an extended late S and/or G2 phase, and are defective in meiosis²⁵. These defects could result from the generation of unresolvable recombination intermediates that trigger checkpoint activation and thereby compromise cell cycle progression. In addition to functioning as an anti-recombinase, Srs2 could conceivably prevent D-loop reversal by removing Rad51 bound to the displaced ssDNA strand. The various activities of Srs2 might be subject to modulation by phosphorylation²⁴.

Other DNA helicase enzymes are known to suppress recombination in eukaryotic cells, including the *S. cerevisiae* Sgs1 protein and the human BLM and WRN proteins, mutated in Bloom's syndrome and Werner's syndrome, respectively. Untimely and aberrant recombination events in Bloom's syndrome and Werner's syndrome cells could contribute to the genomic instability in these cells²⁶. It has been suggested that BLM and WRN proteins control the level of recombination by dissociating recombination intermediates^{5,27}. It will be of interest to test whether Sgs1, BLM and WRN proteins affect the integrity of the hRad51 presynaptic filament, as over-expression of Sgs1 protein can partly suppress some of the defects of *srs2* mutants²⁸. □

Methods

Antibodies and Srs2 purification

Polyclonal antiserum was raised against residues 177–646 of Srs2 fused to glutathione S transferase. Antibodies were purified from the rabbit anti-serum by affinity chromatography on a column containing the antigen crosslinked to cyanogen bromide-activated sepharose 4B matrix (Amersham Biosciences). *SRS2* gene was placed under the T7 promoter in the vector pET11c to yield plasmid pET11c::Srs2, which was introduced into *E. coli* BL21 (DE3). Srs2 expression was induced by isopropyl β-D-thiogalactoside, and extract from 70 l of culture was subjected to precipitation with ammonium sulphate and chromatographic fractionation in columns of Q Sepharose, SP Sepharose, hydroxyapatite and Mono Q. The final Srs2 pool (300 μg at 2 mg ml⁻¹) was nearly homogeneous and stored in small portions at –80 °C.

DNA substrates

The φX circular (+) strand was from New England Biolabs. The φX replicative form I DNA (Gibco-BRL) was linearized by digestion with *Apa*LI. The pBluescript SK(–) replicative form I DNA was prepared as described²⁹. Oligonucleotide D1 has the sequence: 5'-AAATCAATCTAAAGTATATAGAGTAACTTGGTCTGACAGTTACCAATGCTTAA TCAGTGAGGCACCTATCTCAGCGATCTGTCTATTT-3', being complementary to positions 1932–2022 of the pBluescript replicative form I DNA. Oligonucleotide H2 has the sequence: 5'-GTAAGTGTGACACCAAGTTTACTCATATATATCTTTAGATTGATTT-3', being complementary to the first 45 residues of oligonucleotide D1. The two oligonucleotides were 5' end-labelled with [γ-³²P]ATP and purified as described²⁹. The DNA helicase substrate was obtained by hybridizing D1 to radiolabelled H2, as described¹⁴.

Biotinylated dsDNA coupled to magnetic beads

The ends of a 769-base pair fragment derived from digesting φX174 replicative form I DNA with *Apa*LI and *Xho*I were filled in with the Klenow polymerase, using a mixture of dGTP, dTTP, Bio-7-dATP and Bio-11-dCTP (Enzo Diagnostics). The biotinylated DNA

fragment was immobilized on streptavidin-coated magnetic beads (Roche Molecular Biochemicals) to give biotinylated DNA at 40 ng μl⁻¹ packed volume.

Binding of Srs2 to beads containing Rad51

Rad51 and BSA were coupled to Affi-Gel 15 beads (Bio-Rad) at 5 and 12 mg ml⁻¹, respectively¹⁴. To examine Srs2 binding, 3 μg Srs2 was mixed with 7 μl Affi-Rad 51 or Affi-BSA beads in 30 μl PBS (10 mM KH₂PO₄ pH 7.2, 150 mM KCl, 1 mM dithiothreitol (DTT) and 0.01% Igepal) at 4 °C for 30 min. The beads were collected by centrifugation; after the supernatant had been decanted off, the beads were washed twice with 100 μl buffer, then treated for 5 min with 30 μl 2% SDS at 37 °C to elute bound Srs2. The various fractions—10 μl each—were analysed by immunoblotting to determine their Srs2 content.

Yeast two-hybrid assay

RAD51 was cloned into pGAD10, which contains the *GAL4* transcription activation domain, and the resulting plasmid was introduced into the haploid yeast strain PJ69-4a (ref. 30). *SRS2* (residues 1–1174), two C-terminal fragments of *SRS2* (residues 783–1174 and residues 738–998), and *RAD52* were cloned into pGBKT7, which contains the *GAL4* DNA-binding domain; the resulting plasmids were introduced into the haploid yeast strain PJ69-4α (ref. 30). Diploid strains obtained by mating plasmid-bearing PJ69-4a and PJ69-4α haploids were grown on synthetic medium lacking tryptophan and leucine. To select for two-hybrid interactions, which would result in the activation of the *ADE2* and *HIS3* reporter genes, diploid cells were replica-plated on synthetic medium lacking tryptophan, leucine and adenine, and also on synthetic medium lacking tryptophan, leucine and histidine³⁰. Both platings gave identical results. Only the plating on the tryptophan, leucine and adenine dropout medium is shown in Fig. 1e.

DNA helicase assay

Srs2 (35 nM) was incubated at 30 °C for 10 min with the DNA substrate (300 nM nucleotides) in 10 μl buffer H (25 mM Tris-HCl pH 7.5, 2.5 mM MgCl₂, 1 mM DTT, 100 μg ml⁻¹ BSA) containing 2 mM ATP or β-γ-imidoadenosine 5'-phosphate (AMP-PNP) and then analysed¹⁴.

Homologous DNA pairing and strand exchange reaction

Buffer R (35 mM Tris-HCl pH 7.4, 2.0 mM ATP, 2.5 mM MgCl₂, 50 mM KCl, 1 mM DTT, containing an ATP-regenerating system consisting of 20 mM creatine phosphate and 20 μg ml⁻¹ creatine kinase) was used for the reactions, and all the incubation steps were performed at 37 °C. Rad51 (10 μM) was mixed with φX circular (+) strand (30 μM nucleotides) in 30 μl for 5 min, followed by the incorporation of RPA (2 μM) in 1.5 μl and a 3 min incubation. The reaction was completed by adding 3 μl 50 mM spermidine hydrochloride and linear φX dsDNA (30 μM nucleotides) in 3 μl. Portions (4.5 μl) of the reaction mixtures were taken at the indicated times, deproteinized and resolved in agarose gels followed by ethidium bromide staining of the DNA species, as described previously¹⁸. Srs2 was added to the reactions in 0.9 μl at the time of RPA incorporation.

D-loop reaction

Buffer R was used for the D-loop reactions. The radiolabelled oligonucleotide D1 (3 μM nucleotides) was incubated with Rad51 (1 μM) in 22 μl for 5 min at 37 °C, followed by the incorporation of Rad54 (150 nM) in 1 μl and a 2-min incubation at 23 °C. The reaction was initiated by adding pBluescript replicative form I DNA (50 μM base pairs) in 2 μl. The reaction mixtures were incubated at 30 °C, and 5-μl aliquots were withdrawn at the indicated times and processed for electrophoresis as described above. The gels were dried and subjected to phosphorimaging analysis. The percentage of D-loop refers to the quantity of the replicative form substrate that had been converted into D-loop. When present, RPA (200 nM) and Srs2 (40–70 nM) were added to the preassembled Rad51 filament, followed by a 4-min incubation at 37 °C before Rad54 was incorporated. In Fig. 3A, Srs2 (45 nM) was added to the D-loop reactions before Rad54 as above, or 1 min after the incorporation of the duplex substrate. The reactions were terminated after 4 min of incubation.

Topoisomerase-I-linked DNA unwinding assay

Buffer R was used for the reactions and all the incubation steps were performed at 37 °C. Rad51 (4 μM) was incubated for 4 min with pBluescript (–) strand (20 μM nucleotides) in 7.8 μl. Srs2 (40, 60 or 80 nM) and RPA (1 μM) were added in 1 μl, followed by a 4-min incubation. Topologically relaxed φX174 DNA (12.5 μM nucleotides) in 0.8 μl and 2.5 U calf thymus topoisomerase I (Invitrogen) in 0.4 μl storage buffer were then incorporated to complete the reaction. The reaction mixtures were incubated for 8 min and then stopped by adding SDS to 0.5%. In reactions that did not contain the (–) strand, Rad51, with or without RPA (1 μM) and Srs2 (80 nM), was incubated for 8 min with topologically relaxed φX174 DNA and topoisomerase I in a final volume of 10 μl. The reaction mixtures were treated for 10 min with proteinase K (0.5 mg ml⁻¹) before being analysed in 0.9% agarose gels.

Transfer of Rad51 to bead-bound biotinylated dsDNA

M13mp18 circular (+) strand (7.2 μM nucleotides) was incubated for 5 min with Rad51 (2.4 μM) at 37 °C, followed by the addition of Srs2 (30, 60 or 90 nM) in a final volume of 20 μl buffer R containing 50 mM KCl and 0.01% Igepal. After 3 min at 37 °C, 4 μl magnetic beads containing dsDNA were added to the reaction, followed by constant mixing for 5 min at 23 °C. The beads were captured with the Magnetic Particle Separator (Boehringer Mannheim), washed twice with 50 μl buffer, and the bound Rad51 was eluted with 20 μl 1% SDS. The supernatant, which contained unbound Rad51, and the SDS eluate (10 μl each) were analysed by SDS–polyacrylamide-gel electrophoresis (SDS–PAGE).

Electron microscopy

The reactions were performed in buffer R at 37 °C and had a final volume of 12.5 µl. To assemble the Rad51 presynaptic filament, M13mp18 (+) strand (7.2 µM nucleotides) and 1.3 µg Rad51 (2.4 µM) were incubated for 5 min. To test the effects of Srs2 and RPA, these proteins were added to the reaction mixtures containing the preassembled Rad51 presynaptic filament to final concentrations of 60 nM (Srs2) and 350 nM (RPA), followed by a 3-min incubation. In some cases, linear dsDNA (a 5.2-kilobase fragment derived from the pET24 vector) was also added with Srs2 and RPA to 7.2 µM base pairs, followed by a 5-min incubation. For electron microscopy, 3 µl of each reaction mixture was applied to copper grids coated with thin carbon film after glow-discharging the coated grids for 2 min. The grids were washed twice with buffer R and stained for 30 s with 0.75% uranyl formate. After air-drying, the grids were examined with a Philips Tecnai12 electron microscope under low-dose conditions. Images were recorded either with a charge-coupled device camera (Gatan) or on Kodak SO-163 films at × 30,000 magnification and then scanned on a SCAI scanner (Zeiss). The experiments shown in Fig. 4 were each independently repeated three or more times and at least 100 nucleoprotein complexes were examined in each experiment.

Received 3 February; accepted 20 March 2003; doi:10.1038/nature01577.

- Klein, H. L. A radical solution to death. *Nature Genet.* **25**, 132–134 (2000).
- Vaze, M. B. *et al.* Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. *Mol. Cell* **10**, 373–385 (2002).
- Lee, S. K., Johnson, R. E., Yu, S. L., Prakash, L. & Prakash, S. Requirement of yeast SGS1 and SRS2 genes for replication and transcription. *Science* **286**, 2339–2342 (1999).
- Gangloff, S., Soustelle, C. & Fabre, F. Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases. *Nature Genet.* **25**, 192–194 (2000).
- Oakley, T. J. & Hickson, I. D. Defending genome integrity during S-phase: putative roles for RecQ helicases and topoisomerase III. *DNA Repair* **1**, 175–207 (2002).
- Rong, L. & Klein, H. L. Purification and characterization of the SRS2 DNA helicase of the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **268**, 1252–1259 (1993).
- Milne, G. T., Ho, T. & Weaver, D. T. Modulation of *Saccharomyces cerevisiae* DNA double-strand break repair by SRS2 and RAD51. *Genetics* **139**, 1189–1199 (1995).
- Chanet, R., Heude, M., Adjiri, A., Maloel, L. & Fabre, F. Semidominant mutations in the yeast Rad51 protein and their relationships with the Srs2 helicase. *Mol. Cell. Biol.* **16**, 4782–4789 (1996).
- Schild, D. Suppression of a new allele of the yeast RAD52 gene by overexpression of RAD51, mutations in srs2 and ccr4, or mating-type heterozygosity. *Genetics* **140**, 115–127 (1995).
- Sung, P. Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science* **265**, 1241–1243 (1994).
- Sung, P., Trujillo, K. M. & Van Komen, S. Recombination factors of *Saccharomyces cerevisiae*. *Mutat. Res.* **451**, 257–275 (2000).
- Cox, M. M. Recombinational DNA repair of damaged replication forks in *Escherichia coli*: questions. *Annu. Rev. Genet.* **35**, 53–82 (2001).
- Bianco, P. R., Tracy, R. B. & Kowalczykowski, S. C. DNA strand exchange proteins: a biochemical and physical comparison. *Front. Biosci.* **3**, 570–603 (1998).
- Petukhova, G., Stratton, S. & Sung, P. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* **393**, 91–94 (1998).
- Mazin, A. V., Zaitseva, E., Sung, P. & Kowalczykowski, S. C. Tailed duplex DNA is the preferred substrate for Rad51 protein-mediated homologous pairing. *EMBO J.* **19**, 1148–1156 (2000).
- Van Komen, S., Petukhova, G., Sigurdsson, S. & Sung, P. Functional cross-talk among Rad51, Rad54, and replication protein A in heteroduplex DNA joint formation. *J. Biol. Chem.* **277**, 43578–43587 (2002).
- Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. Similarity of the yeast RAD51 filament to the bacterial RecA filament. *Science* **259**, 1896–1899 (1993).
- Sung, P. & Roberson, D. L. DNA strand exchange mediated by a RAD51–ssDNA nucleoprotein filament with polarity opposite to that of RecA. *Cell* **82**, 453–461 (1995).
- Sung, P. Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes Dev.* **11**, 1111–1121 (1997).
- Sugiyama, T., Zaitseva, E. M. & Kowalczykowski, S. C. A single-stranded DNA-binding protein is needed for efficient presynaptic complex formation by the *Saccharomyces cerevisiae* Rad51 protein. *J. Biol. Chem.* **272**, 7940–7945 (1997).
- Veaute, X. *et al.* The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments. *Nature* **423**, 309–312 (2003).
- Aboussekhr, A. *et al.* RADH, a gene of *Saccharomyces cerevisiae* encoding a putative DNA helicase involved in DNA repair. Characteristics of radH mutants and sequence of the gene. *Nucleic Acids Res.* **17**, 7211–7219 (1989).
- Aboussekhr, A., Chanet, R., Adjiri, A. & Fabre, F. Semidominant suppressors of Srs2 helicase mutations of *Saccharomyces cerevisiae* map in the RAD51 gene, whose sequence predicts a protein with similarities to procaryotic RecA proteins. *Mol. Cell. Biol.* **12**, 3224–3234 (1992).
- Liberi, G. *et al.* Srs2 DNA helicase is involved in checkpoint response and its regulation requires a functional Mec1-dependent pathway and Cdk1 activity. *EMBO J.* **19**, 5027–5038 (2000).
- Palladino, F. & Klein, H. L. Analysis of mitotic and meiotic defects in *Saccharomyces cerevisiae* SRS2 DNA helicase mutants. *Genetics* **132**, 23–37 (1992).
- Adams, M. D., McVey, M. & Sekelsky, J. J. *Drosophila* BLM in double-strand break repair by synthesis-dependent strand annealing. *Science* **299**, 265–267 (2003).
- Wu, L., Davies, S. L., Levitt, N. C. & Hickson, I. D. Potential role for the BLM helicase in recombinational repair via a conserved interaction with RAD51. *J. Biol. Chem.* **276**, 19375–19381 (2001).
- Mankouri, H. W., Craig, T. J. & Morgan, A. SGS1 is a multicopy suppressor of srs2: functional overlap between DNA helicases. *Nucleic Acids Res.* **30**, 1103–1113 (2002).
- Petukhova, G., Stratton, S. A. & Sung, P. Single strand DNA binding and annealing activities in the yeast recombination factor Rad59. *J. Biol. Chem.* **274**, 33839–33842 (1999).
- Krejci, L., Damborsky, J., Thomsen, B., Duno, M. & Bendixen, C. Molecular dissection of interactions between Rad51 and members of the recombination-repair group. *Mol. Cell. Biol.* **21**, 966–976 (2001).

Acknowledgements We thank M. Sehorn and K. Trujillo for reading the manuscript. This work was supported by research grants from the NIH (H.K., T.E. and P.S.). S.V.K. was supported in part by a predoctoral fellowship from the US Department of Defense, and Y.L. was supported by a NIH postdoctoral fellowship. The molecular electron microscopy facility at Harvard Medical School was established by a donation from the Giovanni Armeise Harvard Center for Structural Biology, and is maintained through a NIH grant.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.K. (krejci@uthscsa.edu) or P.S. (sung@uthscsa.edu).

The Srs2 helicase prevents recombination by disrupting Rad51 nucleoprotein filaments

Xavier Veaute*, Josette Jeusset†, Christine Soustelle*‡, Stephen C. Kowalczykowski§, Eric Le Cam† & Francis Fabre*

* CEA, DSV, Département de Radiobiologie et Radiopathologie, UMR217 CNRS/CEA, BP6, 92265 Fontenay aux Roses Cedex, France

† Interactions Moléculaires et Cancer, UMR 81126 CNRS/IGR/UPS, Institut Gustave Roussy, Rue Camille Desmoulins, 94805 Villejuif Cedex, France
§ Sections of Microbiology and of Molecular and Cellular Biology, Center for Genetics and Development, University of California, Davis, California 95616-8665, USA

Homologous recombination is a ubiquitous process with key functions in meiotic and vegetative cells for the repair of DNA breaks. It is initiated by the formation of single-stranded DNA on which recombination proteins bind to form a nucleoprotein filament that is active in searching for homology, in the formation of joint molecules and in the exchange of DNA strands¹. This process contributes to genome stability but it is also potentially dangerous to cells if intermediates are formed that cannot be processed normally and thus are toxic or generate genomic rearrangements. Cells must therefore have developed strategies to survey recombination and to prevent the occurrence of such deleterious events. In *Saccharomyces cerevisiae*, genetic data have shown that the Srs2 helicase negatively modulates recombination^{2,3}, and later experiments suggested that it reverses intermediate recombination structures^{4–7}. Here we show that DNA strand exchange mediated *in vitro* by Rad51 is inhibited by Srs2, and that Srs2 disrupts Rad51 filaments formed on single-stranded DNA. These data provide an explanation for the anti-recombinogenic role of Srs2 *in vivo* and highlight a previously unknown mechanism for recombination control.

Several phenotypes (discussed below) conferred by the *srs2* deletion are suppressed by mutations that prevent formation of the Rad51 nucleofilaments^{8,9}. Two hypotheses could explain this suppression: either Srs2 functions in replication and repair to prevent the formation of toxic recombination structures, or Srs2 disrupts dead-end recombination intermediates, possibly formed after the arrest of the replication fork, to allow repair through alternative pathways. This second proposition led us to ask whether purified Srs2 acts on preformed recombination structures.

Srs2 was expressed from a baculovirus vector in which SRS2 was cloned in frame with a histidine tag at its amino terminus. We showed that the protein fusion expressed in yeast fully complements the sensitivity of *srs2*-deleted cells to radiation (data not shown).

‡ Present address: UMR2167 CNRS Centre de Génétique Moléculaire, Bâtiment 26, avenue de la Terrasse, 91198 Gif-sur-Yvette Cedex, France.

After three steps of purification, summarized in Fig. 1a, the protein was nearly homogeneous (Fig. 1b).

Srs2 shares homologies with the bacterial UvrD, Rep³ and PcrA helicases that extend beyond the helicase domains. These helicases belong to the SF-1 superfamily and are believed to translocate on single-stranded DNA (ssDNA), supported by their ssDNA-dependent ATPase activity¹⁰. We tested the ATPase activity of Srs2 with ssDNA or double-stranded DNA (dsDNA) as cofactors. Srs2 was previously shown to possess ATPase activity stimulated by M13mp18 ssDNA¹¹. Because this ssDNA adopts secondary structures with dsDNA character¹², we used (dT)₅₅ ssDNA and a 55-mer blunt-ended dsDNA. The ATPase activity of Srs2 is highly stimulated by the ssDNA, whereas dsDNA has no effect (Fig. 1c). We then tested the helicase activity and found that it unwinds 3' but not 5' end-tailed duplex DNA, indicating a preferential 3' → 5' polarity, as previously shown¹¹. We also found that it does not unwind a 24-mer blunt-ended dsDNA substrate (Fig. 1d). Taken together, these results show an exclusive requirement for ssDNA for the Srs2 activity.

To determine whether Srs2 inhibits recombination *in vitro*, we studied its effect on Rad51-promoted DNA strand exchange reactions involving ΦX174 linear dsDNA and circular ssDNA, using a protocol in which the ssDNA was incubated sequentially with Rad51

and the ssDNA binding factor RPA before addition of the dsDNA (Fig. 2a). In the first experiment, we examined whether Srs2 prevents DNA strand exchange. After formation of the Rad51–ssDNA nucleoprotein filament, linear dsDNA was added in the presence of increasing Srs2 concentrations. Samples were taken after 60 min of incubation. In the absence of Srs2, substantial levels of nicked circular duplex, the final DNA strand exchange product, and joint molecules were obtained (Fig. 2b, lane 1). When increasing amounts of Srs2 were introduced, a marked inhibition of strand transfer occurred and no nicked circular duplex was detected at Srs2 concentrations of 132 nM or higher (Fig. 2b, lanes 2–6, and Fig. 2c). This amount of Srs2 is about 1/100 that of Rad51, indicating that Srs2 acts catalytically. To further investigate the mechanism by which Srs2 inhibits DNA strand exchange, we studied its effect when being added during an ongoing DNA strand exchange reaction. A standard reaction, without Srs2, is shown in Fig. 2d,

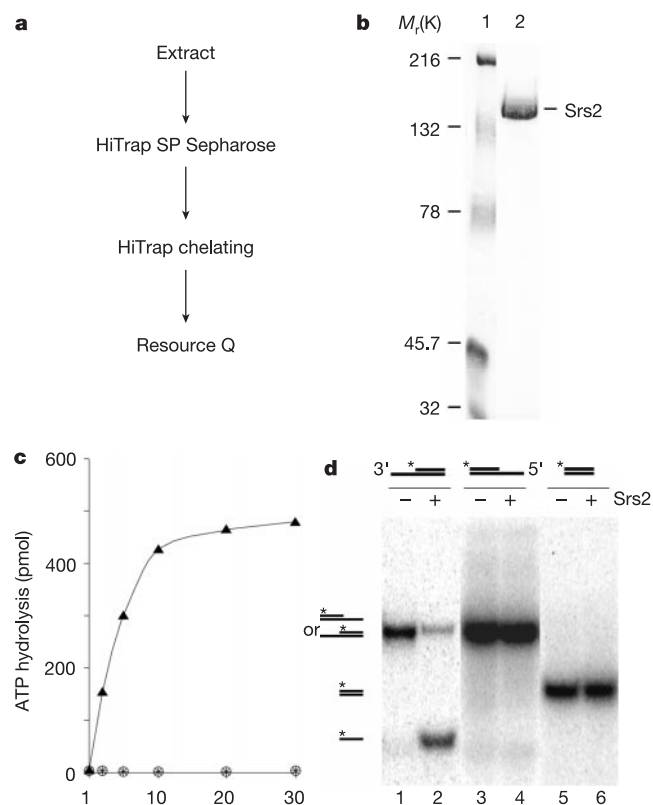


Figure 1 Purification, ATPase and helicase activities of Srs2. **a**, **b**, The Srs2 protein overexpressed in *Spodoptera frugiperda* cells was purified by three chromatographic steps (**a**) to near homogeneity as judged by analysis using 8% SDS–polyacrylamide gel electrophoresis and staining with Coomassie blue (**b**). **c**, Srs2 protein is a ssDNA-dependent but not a dsDNA-dependent ATPase. Purified Srs2 was incubated with [γ -³²P]ATP in the absence of DNA (asterisks) or in the presence of either (dT)₅₅ (triangles) or a 55-mer blunt-ended dsDNA (circles). **d**, Srs2 protein unwinds a γ -³²P-labelled partial duplex DNA with a 3' single-strand tail (lanes 1 and 2) but has no effect on either 5' overhang (lanes 3 and 4) or blunt-ended duplex (lanes 5 and 6).

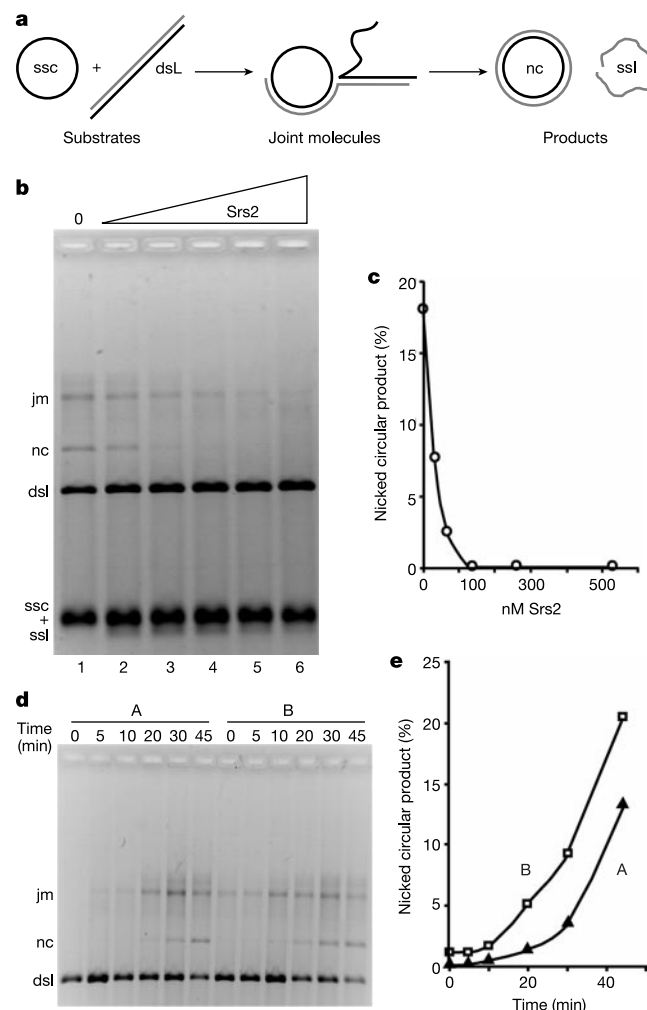


Figure 2 Srs2 inhibits DNA strand exchange catalysed by Rad51. **a**, Scheme of DNA strand exchange reaction. **b**, Increasing amounts of Srs2 were added simultaneously to linear dsDNA in a strand exchange reaction (lanes 1–6 correspond to 0, 33, 66, 132, 264 and 528 nM Srs2, respectively). **c**, Quantification of the reactions shown in **b**. **d**, Panel A: time course of a standard DNA strand exchange. Srs2 (264 nM) was added at various times after initiation (as indicated) and incubation was continued for 60 min. **e**, Quantification of the reactions shown in **d**. Abbreviations: dsl, double-stranded linear DNA; jm, joint molecules; nc, nicked circular double-stranded DNA; ssl, single-stranded linear DNA; ssc, single-stranded circular DNA.

panel A. Samples were recovered at different incubation times. In parallel, and at the corresponding times, Srs2 was introduced at a concentration of 264 nM and incubation was continued for a further 60 min (Fig. 2d, panel B). The comparison of results obtained in these two assays indicates no evidence for dissociation of the Rad51-made DNA joint molecules by Srs2. It might be that the helicase activity of Srs2 is not processive enough to unwind long dsDNA, which would have led to the observable dissociation of joint molecules. In fact, it seems that Srs2 freezes the reaction very quickly after its addition. The quantity of DNA strand exchange products for each time point is slightly higher in the samples to which Srs2 was added and incubated further (Fig. 2e). This is probably due to the time needed for Srs2 to act and block the reaction totally. One hypothesis accounting for this set of results is that Srs2 removes Rad51 from the ssDNA by means of its translocase activity. Elimination of the presynaptic filament would indeed stop the DNA strand exchange reaction.

Electron microscopy was used to reveal the possible disruptive effect of Srs2 on Rad51 nucleofilaments. Rad51 was first incubated with Φ X174 viral (+) strand for 20 min to assemble nucleoprotein filaments. All of the observed molecules were similar to that shown in Fig. 3a. When Srs2 (105 nM) was then added to the reaction and the incubation was continued for 2 min, Rad51 nucleoprotein filaments were disrupted to various extents. Of 532 molecules counted, 15% had partial nucleofilaments dismantled at different locations (Fig. 3b), whereas the others were revealed only by short nucleofilament structures. After 10 min of incubation with Srs2, no molecules were detected. The presence of naked DNA molecules was revealed by subsequent addition of the T4 gene 32 ssDNA-binding protein (Fig. 3c). The 452 molecules observed were all bound by T4 gene 32 protein, as shown in Fig. 3d, and were found at a concentration similar to that of the Rad51 nucleofilaments. T4 gene 32 protein alone does not displace Rad51 bound to ssDNA (580 molecules observed; data not shown). These experiments

indicate that Srs2 displaces Rad51 from ssDNA. Last, we found that Srs2 does not disrupt Rad51 filaments formed on Φ X174 linear dsDNA (Fig. 3e, f). Thus, the removal of the Rad51 nucleofilaments by Srs2 occurs only on ssDNA, by a process that is probably linked to the translocation of Srs2 along ssDNA. This activity of Srs2 has also been recently shown by Sung *et al.*¹³.

We next asked whether Srs2 is also active on RecA ssDNA nucleoprotein filaments. Srs2 was found to inhibit strand exchange and, by electron microscopy, to disrupt RecA–ssDNA nucleoprotein filaments (data not shown). Thus, Srs2 might have specificity for nucleoprotein filaments formed with the RecA/Rad51 class of recombinases. The fact that *SRS2*-deleted cells have strong phenotypes explainable by an abnormal persistence of Rad51 nucleofilaments most probably indicates that the disruption of nucleofilaments is an important role of Srs2 that is not replaced *in vivo* by other helicases. We now discuss some of these phenotypes.

The *RAD51*-dependent ultraviolet sensitivity of *srs2* cells¹⁴ is explained by the binding of recombination proteins on ssDNA, formed before and/or after replication, generating toxic DNA structures if not processed by Srs2. Removal of the recombination proteins would be required to allow translesion synthesis. A role for Srs2 was also shown in cells expressing leaky alleles of *RAD51* and *RAD52*. These mutants are sensitive to ionizing radiation, a phenotype largely suppressed by the *SRS2* deletion^{4–7}. In these cases the sensitivity of the mutants would be due to the disruption of the nucleofilament poisoned by the leaky proteins, thus preventing DSB repair. In the absence of Srs2, the leakiness of the system would permit repair. Similarly, we found that the cold sensitivity of *rad55*- or *rad57*-deleted cells to radiation¹⁵ is suppressed by *srs2* (F.F. and X.V., unpublished data), indicating that, in those mutants, filaments are formed but disrupted by Srs2. It is important to note that Rad55 and Rad57 were absent from our assays *in vitro*, and it could be that the observed Srs2 activity *in vivo* relates to this absence.

The *srs2* mutant is also deficient in adaptation (escape from checkpoint arrest when cells suffer irreparable DSBs) and in recovery (the resumption of cell divisions after checkpoint arrest and completion of repair)¹⁶. It was proposed that Srs2 wipes out the checkpoint proteins from the DNA complex, eliminating the checkpoint signal. However, on the basis of the observation that the *srs2* defect in recovery is alleviated by *rad51* (ref. 16), the possibility that the role of Srs2 in recovery is to eliminate persistent Rad51 nucleofilaments that would otherwise trigger checkpoints is not totally excluded.

Last, Srs2 is essential in different mutant contexts, owing to the formation of toxic recombination intermediates. The synthetic lethality of *srs2* and *rad54* (ref. 17) is suppressed by *rad51* (refs 6, 18). The inactive Rad51 ssDNA nucleofilaments formed in *rad54* cells¹⁹ would be toxic if not disrupted by Srs2. The *srs2 sgs1* growth defect also depends on the initiation of recombination²⁰, indicating that in *sgs1* cells potentially toxic recombination intermediates are formed, which are substrates for Srs2. The role of Sgs1 in recombination or/and replication is not yet clear⁸. Interestingly, Sgs1 is a helicase of the recQ family, which has orthologues in human cells that are responsible for the Bloom, Werner and a subset of Rothmund–Thomson syndromes, characterized by genomic instability and a predisposition to cancer²¹. These phenotypes might be related to unrestrained initiation of recombination that can lead to chromosomal instability. It was indeed shown that Wrn cellular phenotypes depend on Rad51-promoted recombination²². Our results should help us to understand the molecular basis of these syndromes.

Genetic data show that Srs2 is triggered by DNA lesions and by mutations affecting the structure and/or the activity of Rad51 nucleofilaments. Our finding supports the view that Srs2 disrupts nucleofilaments essentially when subsequent recombination steps are impaired. A close association between Srs2 and the recombination machinery is therefore predicted. □

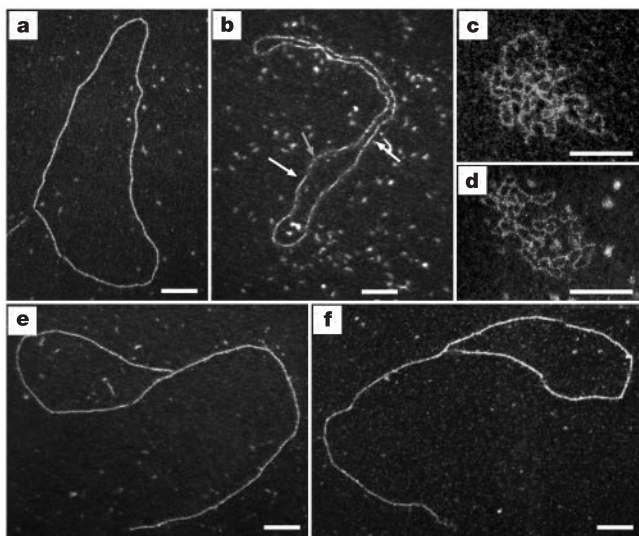


Figure 3 Disruption of Rad51 presynaptic filament by Srs2 as examined by electron microscopy. **a**, Rad51–ssDNA nucleoprotein filaments. **b**, Preformed Rad51–ssDNA filaments were incubated for 2 min with 105 nM Srs2. White arrows show intact Rad51 nucleoprotein filament and the grey arrow indicates regions of destabilized filament. **c**, Φ X174 viral (+) strand covered by T4 gene 32 protein. **d**, Preformed Rad51–ssDNA complexes were incubated for 10 min with 105 nM Srs2. T4 gene 32 protein was then added and incubation was continued for 10 min. **e**, Rad51–dsDNA nucleoprotein filaments. **f**, Rad51–dsDNA nucleoprotein filaments incubated for 10 min with Srs2 protein. Scale bars, 200 nm.

Methods

Construction of recombinant baculovirus

Plasmid p14HB was constructed by deletion of the *Hind*III fragment of p14 (ref. 3) followed by the introduction of a new *Bam*HI site upstream of the *SRS2* open reading frame. The *Bam*HI–*Eco*RI fragment of p14HB containing the *SRS2* gene was cloned into pBacPAK–His1. Recombinant baculovirus was constructed as recommended by the manufacturer (Invitrogen). The recombinant baculovirus encodes the Srs2 protein with 17 additional amino acid residues (MGHHHHHHVVDKLGSM) fused to its N-terminal methionine.

Purification of the recombinant Srs2 protein

Sf9 (*Spodoptera frugiperda*) cells were infected with the recombinant baculovirus at a multiplicity of infection of 10. After incubation for 50 h, cells were harvested, resuspended at 3×10^7 cells ml⁻¹ in 50 mM phosphate buffer pH 7.2, 130 mM NaCl, 1 mM 2-mercaptoethanol, Triton X-100 $\times 1$, 1 mM 4-(2-aminoethyl)benzenesulphonyl fluoride, 2 mM benzaminide, 2 μ M pepstatin, 5 μ g ml⁻¹ leupeptin and disrupted by sonication. The extract was cleared by centrifugation at 40,000 r.p.m. for 1 h. After dilution to achieve a final concentration of 100 mM NaCl and 10% glycerol, it was applied on a HiTrap SP Sepharose HP column (Amersham Biosciences) equilibrated with buffer A (50 mM phosphate buffer pH 7.2, 100 mM NaCl, 1 mM 2-mercaptoethanol and 10% glycerol). The column was eluted with a 60-ml linear gradient of 100–1,000 mM NaCl in buffer A. Fractions containing Srs2 were pooled and loaded on a HiTrap chelating column (Amersham Biosciences) equilibrated with 50 mM phosphate buffer pH 7.2, 500 mM NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol and 10% glycerol. The column was then washed with buffer B (20 mM phosphate buffer pH 7.8, 50 mM NaCl, 20 mM imidazole, 1 mM 2-mercaptoethanol and 10% glycerol). After elution with a 60-ml linear gradient of 20–400 mM imidazole in buffer B, fractions containing Srs2 were pooled and applied to a Resource Q column (Amersham Biosciences) equilibrated with buffer B without imidazole. The column was eluted with a 10-ml linear gradient of 50–600 mM NaCl. Highly purified Srs2 protein eluted at ~200 mM NaCl and was stored at –75 °C. Identity of the Srs2 protein was verified by matrix-assisted laser desorption ionization–time-of-flight mass spectrometry (data not shown).

ATP hydrolysis assay

The Srs2 protein (6 nM) was incubated at 37 °C with DNA cofactor (4.53 μ M nucleotides), 40 μ M [γ -³²P]ATP in 25 μ l reaction buffer (50 mM Tris–HCl pH 7.6, 100 mM NaCl, 7 mM MgCl₂, 5 mM dithiothreitol, 250 μ g ml⁻¹ BSA) and stopped by adding 20 mM EDTA pH 8. The level of ATP hydrolysis was measured as described²³.

Helicase assay

The substrates used to determine the directionality of the Srs2 activity consisted of partial duplex DNA annealed over 24 nucleotides with a single-strand tail of 21 nucleotides. Polarity 3' → 5' was checked with oligonucleotide P2 (5'-ACCTACCTCTGATTGCGAATCTTC-3') annealed with oligonucleotide P1 (5'-TGAAGGTTTCGAATCAGAGGTAGG TGCCCGGCCGATAAATGCGCGTATTCCTGGT-3'). The opposite polarity was checked with oligonucleotide P3 (5'-ACCAGGAATACGCGCATTTATCCG-3') annealed with P1. Blunt-ended duplexes were obtained by annealing P2 with the complementary oligonucleotide (P4). P2 and P3 were 5' end-labelled with [γ -³²P]ATP and polynucleotide kinase. DNA substrate (20 fmol) was incubated at 30 °C with 121 nM Srs2 in 50 mM Tris–HCl pH 7.6, 7 mM MgCl₂, 2 mM ATP, 5 mM dithiothreitol and 25 μ g ml⁻¹ BSA for 30 min. The reaction products were resolved by electrophoresis on a 12% non-denaturing polyacrylamide gel in Tris-borate-EDTA buffer.

DNA strand exchange reaction

Reactions (10.5 μ l) contained 42 mM MOPS pH 7.2, 3 mM magnesium acetate, 1 mM dithiothreitol, 20 mM NaCl, 25 μ g ml⁻¹ BSA, 2.5 mM ATP, 33 μ M (nucleotides) Φ X174 viral (+) strand and 11 μ M Rad51 protein. After incubation at 37 °C for 15 min, 1.1 μ M RPA was added and incubation was continued for 5 min. Strand exchange was started by the addition of 1 μ l (33 μ M nucleotide final concentration) of *Pst*I-linearized Φ X174 dsDNA, 4 mM spermidine and variable amounts of Srs2 protein. After incubation, the samples were deproteinized and analysed by electrophoresis in 0.8% agarose gel in Tris-acetate-EDTA buffer. The gels were stained with ethidium bromide and quantified with ImageQuant software.

Electron microscopy

Rad51 filaments on ssDNA or dsDNA were formed by incubation of 6.48 μ M Rad51 protein in DNA strand exchange buffer without BSA with 11 μ M (nucleotides) Φ X174 viral (+) strand or *Pst*I-linearized Φ X174 replicative-form DNA for 20 min at 37 °C. Srs2 (105 nM final concentration) was then added for 10 min at 37 °C. To detect the presence of naked Φ X174 viral (+) strand after the action of Srs2, T4 gene 32 protein (0.65 μ M) was

subsequently added for 10 min. The reaction mixtures were diluted 1:50 in 10 mM Tris–HCl pH 7.5, 50 mM NaCl and 5 mM MgCl₂ without any chemical fixation, and analysed by electron microscopy as described previously²⁴.

Received 3 February; accepted 17 March 2003; doi:10.1038/nature01585.

1. Sung, P., Trujillo, K. M. & Van Komen, S. Recombination factors of *Saccharomyces cerevisiae*. *Mutat. Res.* **451**, 257–275 (2000).
2. Aguilera, A. & Klein, H. L. Genetic control of intrachromosomal recombination in *Saccharomyces cerevisiae*. I. Isolation and genetic characterization of hyper-recombination mutations. *Genetics* **119**, 779–790 (1988).
3. Aboussekhra, A. *et al.* RADH, a gene of *Saccharomyces cerevisiae* encoding a putative DNA helicase involved in DNA repair. Characteristics of *radH* mutants and sequence of the gene. *Nucleic Acids Res.* **17**, 7211–7219 (1989).
4. Kaytor, M. D., Nguyen, M. & Livingston, D. M. The complexity of the interaction between RAD52 and SRS2. *Genetics* **140**, 1441–1442 (1995).
5. Milne, G. T., Ho, T. & Weaver, D. T. Modulation of *Saccharomyces cerevisiae* DNA double-strand break repair by SRS2 and RAD51. *Genetics* **139**, 1189–1199 (1995).
6. Schild, D. Suppression of a new allele of the yeast RAD52 gene by overexpression of RAD51, mutations in *srs2* and *crr4*, or mating-type heterozygosity. *Genetics* **140**, 115–127 (1995).
7. Chanet, R., Heude, M., Adjiri, A., Maloel, L. & Fabre, F. Semidominant mutations in the yeast Rad51 protein and their relationships with the Srs2 helicase. *Mol. Cell. Biol.* **16**, 4782–4789 (1996).
8. Fabre, F., Chan, A., Heyer, W. D. & Gangloff, S. Alternate pathways involving Sgs1/Top3, Mus81/Mms4, and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication. *Proc. Natl. Acad. Sci. USA* **99**, 1687–1692 (2002).
9. Klein, H. L. Mutations in recombinational repair and in checkpoint control genes suppress the lethal combination of *srs2Δ* with other DNA repair genes in *Saccharomyces cerevisiae*. *Genetics* **157**, 557–565 (2001).
10. Singleton, M. R. & Wigley, D. B. Modularity and specialization in superfamily 1 and 2 helicases. *J. Bacteriol.* **184**, 1819–1826 (2002).
11. Rong, L. & Klein, H. L. Purification and characterization of the Srs2 DNA helicase of the yeast *Saccharomyces cerevisiae*. *J. Biol. Chem.* **268**, 1252–1259 (1993).
12. Petukhova, G., Stratton, S. & Sung, P. Catalysis of homologous DNA pairing by yeast Rad51 and Rad54 proteins. *Nature* **393**, 91–94 (1998).
13. Krejci, L. *et al.* DNA helicase Srs2 disrupts formation of the Rad51 presynaptic filament. *Nature* **423**, 305–309 (2003).
14. Aboussekhra, A., Chanet, R., Adjiri, A. & Fabre, F. Semidominant suppressors of Srs2 helicase mutations of *Saccharomyces cerevisiae* map in the RAD51 gene, whose sequence predicts a protein with similarities to prokaryotic RecA proteins. *Mol. Cell. Biol.* **12**, 3224–3234 (1992).
15. Lovett, S. T. & Mortimer, R. K. Characterization of null mutants of the RAD55 gene of *Saccharomyces cerevisiae*: effects of temperature, osmotic strength and mating type. *Genetics* **116**, 547–553 (1987).
16. Vaze, M. *et al.* Recovery from checkpoint-mediated arrest after repair of a double-strand break requires Srs2 helicase. *Mol. Cell* **10**, 373 (2002).
17. Palladino, F. & Klein, H. L. Analysis of mitotic and meiotic defects in *Saccharomyces cerevisiae* SRS2 DNA helicase mutants. *Genetics* **132**, 23–37 (1992).
18. Heude, M., Chanet, R. & Fabre, F. Regulation of the *Saccharomyces cerevisiae* Srs2 helicase during the mitotic cell cycle, meiosis and after irradiation. *Mol. Gen. Genet.* **248**, 59–68 (1995).
19. Van Komen, S., Petukhova, G., Sigurdsson, S., Stratton, S. & Sung, P. Superhelicity-driven homologous DNA pairing by yeast recombination factors Rad51 and Rad54. *Mol. Cell* **6**, 563–572 (2000).
20. Gangloff, S., Soustelle, C. & Fabre, F. Homologous recombination is responsible for cell death in the absence of the Sgs1 and Srs2 helicases. *Nature Genet.* **25**, 192–194 (2000).
21. van Brabant, A. J., Stan, R. & Ellis, N. A. DNA helicases, genomic instability, and human genetic disease. *Annu. Rev. Genomics Hum. Genet.* **1**, 409–459 (2000).
22. Saintigny, Y., Makienco, K., Swanson, C., Emond, M. J. & Monnat, R. J. Jr Homologous recombination resolution defect in Werner syndrome. *Mol. Cell. Biol.* **22**, 6971–6978 (2002).
23. Iggo, R. D. & Lane, D. P. Nuclear protein p68 is an RNA-dependent ATPase. *EMBO J.* **8**, 1827–1831 (1989).
24. Beloin, C. *et al.* Contribution of DNA conformation and topology in right-handed DNA wrapping by the *Bacillus subtilis* LrpC protein. *J. Biol. Chem.* **278**, 5333–5342 (2003).

Acknowledgements We thank L. Miccoli, D. Biard and J. Angulo for advice on recombinant baculovirus preparation and protein purification. P. Bertrand and C. Auvin for confirming the identity of Srs2 by MALDI–TOF mass spectrometry, and S. Gangloff, N. Kantake, L. Leloup, L. Maloel, J. New and T. Robert for comments and discussions. This work was supported by the Commissariat à l’Énergie Atomique, the Centre National de la Recherche Scientifique and Électricité de France.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to X.V. (xavier.veaute@cea.fr) or F.F. (francis.fabre@cea.fr).

Chromatography developments

Acronyms abound in two pages devoted to GC, TLC, GPC, HPLC etc.

761 Compact IC system PIC 2

Metrohm www.metrohm.co.uk

Good things come in small packages

Researchers who are considering moving to ion chromatography (IC) may be interested in the Metrohm 761 IC system, which packs a full set of IC components into a minimum amount of bench space — about half the size of some alternative machines. It has an injection valve for individual injections or for use with a sample changer, a double piston pump with a flow range of 0.2–2.5 ml per minute and a maximum pressure of 25 MPa (250 bar), and a pulsation damper to protect columns from damage. In addition, a shielded column chamber for anion columns, with or without suppression, a suppressor module, a two-channel peristaltic pump, a conductivity detector and a high-resolution AD converter are all contained within the housing. The system's software can be run on a Windows system and requires 20 MB of free memory, a Pentium processor or higher and 64 MB of internal memory, with Windows 95 or NT.



Half measures: a 'compact' from Metrohm.

tenance, warning and error messages shown in words rather than code, in any of seven languages. Other features include a liquid autosampler, programmable pneumatic control, isolation and enhanced solvent purge modes and large volume injection. The mass spectrometer uses simultaneous selected ion and full ion scanning technology to optimize accuracy and spectral fidelity.

Chemicals catalogue

Fisher Scientific www.fisherchemicals.com

Of consuming interest

Fisher Scientific's 2003 catalogue includes a section devoted to chromatography, offering a wide range of solids, solvents and chromatography consumables. New products include thin-layer chromatography plates and silicas for flash/low-pressure chromatography, a range of irregular silicas, ultra-pure acids and water, and new or reformulated products from the Aqualine pyridine-free reagents range. Advanced high-performance liquid chromatography (HPLC) reagents that demonstrate minimal baseline drift under gradient conditions are available for use in HPLC method development. Chemicals comply with BP, EP and USP standards for use in process development applications. The catalogue is available on the company's website.

Clarus 500 GC/MS

PerkinElmer
www.perkinelmerinstruments.co.uk

It speaks your language

The Clarus 500 gas chromatograph features a full-colour touch screen that displays a real-time chromatogram from either channel A or channel B, allowing users to monitor a current analysis on screen. Pictorial icons are used to display system status, with main-

enhanced to support FDA 21 CFR Part 11 compliance, and includes provision for application-to-application communications, electronic signatures, triple-digit levels of password protection and full audit trail implementation. A runtime automation feature controls automatic data collection and reporting.

CLASS-VP v7.2

Shimadzu www.shimadzu.com

Advanced compliance with CFR Part 11

Shimadzu's CLASS-VP v7.2 client/server chromatography data system has been upgraded to ensure compliance with the US FDA's 21 CFR Part 11 requirements. These include a signing authority hierarchy to enforce proper application of electronic signatures, optional comments for e-signatures, thin client interaction to simplify validation of large software deployments and expanded system administration activity logs and audit trails. The software can be used with a wide range of high-performance liquid chromatography and gas chromatography systems, including digitally configured injection and detector modules, and is aimed at users in pharmaceutical laboratories.

Validation toolkit for Atlas 2002 CDS

Thermo LabSystems
www.thermolabsystems.com

Test-drive it first

An automated test driver has been added to the validation toolkit for version 2002 of the Atlas chromatography data system (CDS). The kit also includes automated test scripts, with manual versions of the tests in Microsoft Word format, a visual test source code, functional specifications to allow users to tailor testing to their own needs, and a test approach/functional specification matrix.

Thermo LabSystems - Atlas Automated Test Driver	
File	Tools Execution Help
Options	Check Results Pre and Post Editor
Test Suite	Description
Boundary	Tests that boundary test fields in Atlas.
Calibration and Quantitation	Tests that check the calibration and quantitation functions.
Change History	Tests that check the Change history.
ChromView	Tests that check the Chromatograms.
Configuration Manager	Tests Configuration Manager.
Develop Integration	
D5K	Tests that check the control of A2D devices.
Exporting	Exporting tests exporting Atlas data.
File System	Tests that check the File System.
Group ID	Tests that check the Group ID functionality in Atlas.
Help	Tests Atlas Help menu.
Identification	Tests that check the Identification functionality in Atlas.
InstControl	Tests that check HP1050, HP1100 and HP5890 instruments.
Instrument	Tests Instrument Manager.
Instrument Hiding	Test instrument hiding.
Multichrom Import	Test Multichrom imports of Atlas.
Peak Detection (Integration)	Tests that check the Peak Detection/Integration function.

Data screen: Atlas for data validation.

Summit P680

Dionex www.dionex.com

Don't put a damper on it

Dionex offers four models in its Summit range of P680 high-performance liquid chromatography pumps: an isocratic version, low- and high-pressure gradient models, and a semi-preparative pump. The company's SmartFlow technology eliminates the need for manual adjustment of compressibility settings. A pulse damper is not required. Interchangeable mixer options accommodate special applications, allowing the user to optimize the gradient pumps for smallest delay volume (<70 ml for the high-pressure gradient model) or for complete mixing of poorly miscible eluents.

IDM-LimsLink v3.1/IDM-LimsLinkCDS v3.1

Applied Biosystems
www.europe.appliedbiosystems.com

Make sure your connections are secure

IDM-LimsLink v3.1 software can be used to link Applied Biosystems' SQL LIMS software with most available chromatography systems. Security and accountability has been

Tests for validation of Thermo's DataServer and A2D instrument data acquisition and control devices, which are supplied with the Atlas CDS, are also included.

Model 301 detector

ESA Analytical www.esainc.com
Sensitive where it counts

The Model 301 evaporative light scattering detector from ESA Analytical expands the versatility of gradient high-performance liquid chromatography to compounds that are not detectable by ultraviolet or fluorescence methods. The technique is fully compatible with aqueous mobile phases, even at high or microbore flow rates. The company claims that the Model 301 provides universal detection of non-volatile compounds with sensitivity that is better than that achieved with RI or low-wavelength UV, and is particularly suitable for detection of weakly chromophoric compounds, including lipids, triglycerides, carbohydrates, surfactants and polymers. Direct assessment of relative mass may be performed without standards owing to the detector's insensitivity to temperature variation and consistency in measurement of relative response factors.

OmniSEC v2.0

Viscotek www.viscotek.com
Multi-detection GPC analyses

Version 2.0 of the OmniSEC gel permeation chromatography/size exclusion chromatography (GPC/SEC) program from Viscotek incorporates a suite of new features designed to maximize information extraction in polymer and protein characterization experiments. The software calculates conventional

and universal calibration data and performs copolymer analysis, handling the full range of GPC for any sample type or calculation technique. A peak detection algorithm provides automatic multi-detection GPC analyses in production and quality control environments. The program performs multiple peak detection and calculation of molecular weight averages and distribution plots for one or more peaks per sample. Up to six channels of data can be imported and analysed simultaneously.

Astec-3 eluent degasser

Advanced Separation Technologies
www.astecusa.com

Just passing through

The Astec-3 triple-channel degasser for liquid chromatography removes the dissolved gas in high-performance liquid chromatography (HPLC) eluents. Eluent flows through the degassing device, which is made of a macromolecular membrane. Since the dissolved gas in the eluent has a smaller molecular size and stronger affinity with the membrane than the eluent, it can pass through the film selectively and efficiently. The degasser has a small internal volume for rapid solvent displacement and can be used with any HPLC pump up to 20 ml per minute.

Nebula 215 LC/MS autopurification system

Gilson www.gilson.com
Add samples on the fly

Gilson's 215 analysis system has been integrated with Thermo Finnigan's compact Surveyor MSQ single quadrupole mass spectrometry detector for classical high-performance liquid chromatography to produce the NEBULA LC/MS autopurification system. This versatile single-probe injector and fraction collector can perform both functions on the same bed, and samples can be added at any time during operation. Fractions can be collected based on either MS or ultraviolet, with full-scan MS data collected on every sample. Selectivity of MS-based fraction collection minimizes 'non-hit' fractions. Chromatographic trace is generated from the summed extracted ion chro-



Gas guzzler: Astec-3 tackles eluent gas.

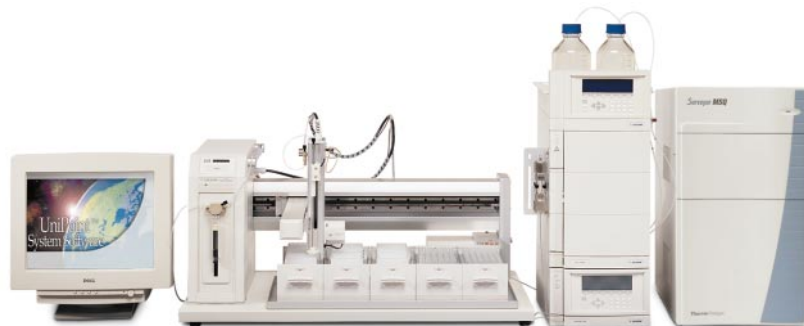
matogram of up to seven masses. Real-time traces from both MS and UV/VIS detectors allow on-screen comparison of results. The system is suitable for work in combinatorial chemistry as well as pharmaceutical and biochemical applications.

Software demo CD-ROM

ChemSW www.chemsw.com
Give it a spin, for free

ChemSW is offering a free CD-ROM containing demonstrations of more than 30 of the company's software programs for chemistry laboratories, along with a list of over 40 books concerned with mass spectrometry and gas chromatography. The collection includes demonstration versions of Mass Spec Calculator Pro, NIST/EPA/NIH Mass Spectral Search Library, Mass Spec Tools, Interactive Training Modules, GC and MS graph editing, and protein chemistry tools. The CD can be ordered via email at info@chemsw.com.

These notes are compiled in the Nature office from information provided by the manufacturers. Material intended for publication can be sent to newproducts@nature.com.



Chromatography line-up: Gilson's 215 analysis system and peripherals.

nature insight

bone and cartilage



Cover illustration

Inset, lumbar vertebrae of a newborn mouse stained with alcian blue (cartilage) and alizarin red (bone). Background, chondrocytes in the phalange of a newborn mouse; blue coloration signals expression of 'digit genes' of the *HoxD* gene cluster. (Courtesy of Marie Kmita, Jozsef Zakany and Denis Duboule.)

Bone and cartilage comprise the skeleton — a finely patterned structure that provides mobility, protection of vital organs, and housing of the bone marrow. Because bones are the most inert and enduring of our mortal remains, it is surprising what dynamic and highly tuned organs they are during life, as they must simultaneously balance growth to achieve strength and resilience, and repair without overgrowth.

This balance is achieved by bone remodelling — an interplay of mineral deposition and resorption by specialized bone cells, which is constantly occurring on a microscopic scale throughout our bodies. An imbalance in remodelling is what leads to the debilitating loss of bone mass called osteoporosis, a condition that is estimated to afflict 200 million women world wide, and often leads to immobility and death. This and other diseases of bone and cartilage are responsible for a large portion of healthcare expenditures in developed countries — US\$14 billion is spent annually on treating osteoporotic fractures in the United States alone.

These statistics emphasize the need for a better understanding of both the fundamentals of skeletal biology and the pathologies associated with it, which in addition to osteoporosis, include birth defects, genetic diseases, and rheumatoid and osteoarthritis. Indeed, as human life is prolonged by biomedical advances, the weak link in our health and longevity may become the strength of our bones and mobility of our joints.

This month's Insight presents a collection of articles that explore some key elements of skeletal biology, including how skeletal structures are patterned and bones and cartilage develop, the influence of genetics over human skeletal biology, the signalling pathways and transcription factors that control bone mass, and the pathogenic mechanisms of rheumatoid arthritis.

We are pleased to acknowledge the financial support of the National Institutes of Health in producing this Insight. As always, *Nature* carries sole responsibility for all editorial content and peer-review.

Natalie DeWitt Senior Editor

overview:

- 316 The complexities of skeletal biology**
G. Karsenty

review articles:

- 319 Deciphering skeletal patterning: clues from the limb**
F. V. Mariani & G. R. Martin
- 326 Cranial skeletal biology**
J. A. Helms & R. A. Schneider
- 332 Developmental regulation of the growth plate**
H. M. Kronenberg
- 337 Osteoclast differentiation and activation**
W. J. Boyle, W. S. Simonet & D. L. Lacey
- 343 The genetic basis for skeletal diseases**
E. Zelzer & B. R. Olsen
- 349 Control of osteoblast function and regulation of bone mass**
S. Harada & G. A. Rodan
- 356 Evolving concepts of rheumatoid arthritis**
G. S. Firestein

Editor, *Nature*: Philip Campbell
Insights Editors: Lesley Anson
Ursula Weiss
Editorial Assistant: Simon Gibson

Art Director: Majo Xeridat
Layouts: Clifford Saunders
Diagrams: Ann Thomson
Nell Greenhill

Production Editor: Simon Gribbin
Production: Sue Gray
Marketing: Stephanie Liu
Sponsorship: Mark Greene

The complexities of skeletal biology

Gerard Karsenty

Baylor College of Medicine, Bone Disease Program of Texas, One Baylor Plaza, Houston, Texas 77030, USA (e-mail: karsenty@bcm.tmc.edu)

For a long time, the skeleton was seen as an amorphous tissue of little biological interest. But such a view ignored the large number of genetic and degenerative diseases affecting this organ. Over the past 15 years, molecular and genetic studies have modified our understanding of skeletal biology. By so doing this progress has affected our understanding of diseases and suggested in many instances new therapeutic opportunities.

The skeleton is an organ of unappreciated complexities. These complexities affect patterning and cell differentiation during development, and physiology and pathology postnatally. Its peculiar distribution in more than 220 locations in the human body and the extreme conservation of the function of genes affecting skeletal development and physiology between mouse and human explains why our ability to generate genetically modified mice has so profoundly transformed skeletal biology. Probably no other organ in vertebrate biology has benefited as much from our ability to generate, at will, mutant mouse strains.

The skeleton is made of two tissues (cartilage and bone), three cell types (chondrocytes, osteoblasts and osteoclasts) and more than 200 different skeletal elements spread out throughout the body. These properties raise questions about the location and shape of skeletal elements as well as allocation of cell lineage. Beyond development, the skeleton has to fulfil a series of functions about which we have little understanding in molecular terms, but which are of critical importance as they are often affected in common degenerative diseases. Among these functions one can cite the molecular mechanisms determining the extent of longitudinal growth, the spatial restriction of extracellular matrix mineralization to bone, and the maintenance of a constant bone mass.

Although still incomplete, our molecular understanding of skeletal physiology has advanced considerably by analysing skeletal cells themselves as well as how hormones and the nervous system affects their proliferation and function. As described in this timely issue, significant progress has been made in addressing the developmental biology, physiology and pathology of the skeleton. This progress in turn has raised a series of questions relative to each of the different aspects of skeletal biology.

Development of the skeleton

The aspect of skeletal biology that first received attention from molecular geneticists was developmental biology. One reason for this, and a key feature of the skeleton, is that the shape of skeletal elements varies greatly from one location to another in the body. Thus, the study of patterning of skeletal elements or of a group of skeletal elements such as limbs or the skull has been an object of study of embryologists even before the era of molecular biology. The earlier studies were performed mostly in chicken, whereas the more recent molecular studies were done using either chicken or mouse models¹⁻³. Another factor favouring research on developmental biology of the skeleton in the past 15 years is the emerging notion that regulatory genes are often conserved throughout evolution. And even if the function of a given gene is not

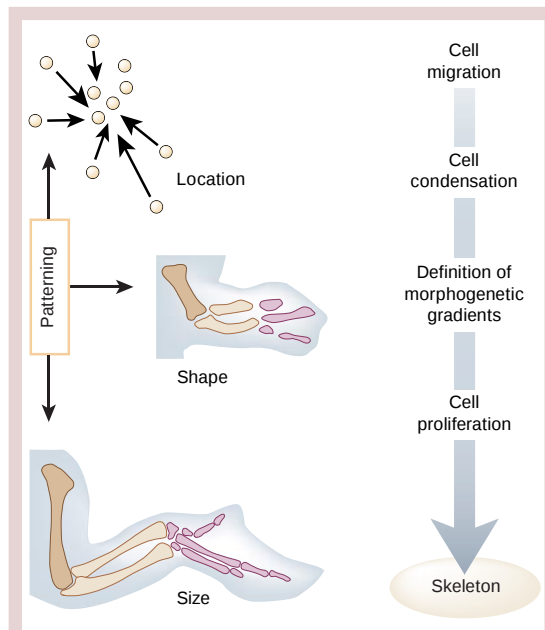


Figure 1 Key issues in skeletal development. The three main issues are: what triggers mesenchymal cells to migrate to a given location and to aggregate in the shape of a future skeletal element; how are morphogenetic gradients defined and regulated to determine location and shape of skeletal elements; and how do skeletal cells differentiate and proliferate?

conserved from lower organisms to mice, the genetic pathways in which it resides very often is. In many instances this has accelerated the elucidation of the molecular bases for a given skeletal phenotype observed in mouse embryos.

The first event studied during development of the skeleton was patterning — the size, shape, number and arrangement of bones of the limbs and of the skull (see reviews in this issue by Mariani and Martin, page 319, and Helms and Schneider, page 326). The embryonic limb emerges from the lateral-plate mesoderm as a bud of mesenchymal cells covered by a layer of ectoderm. These anatomically distinct zones of the developing limb bud — the apical ectodermal ridge, the zone of polarizing activity and the progress zone — concur to establish over time the proximal–distal, anterior–to–posterior and dorsal–to–ventral axes of the limb. Many of the genes that are expressed in each area and that control limb patterning have been identified and their function characterized⁴.

The principal debate in this field now centres on how cells in the progress zone acquire positional information. One model proposes that acquisition occurs progressively

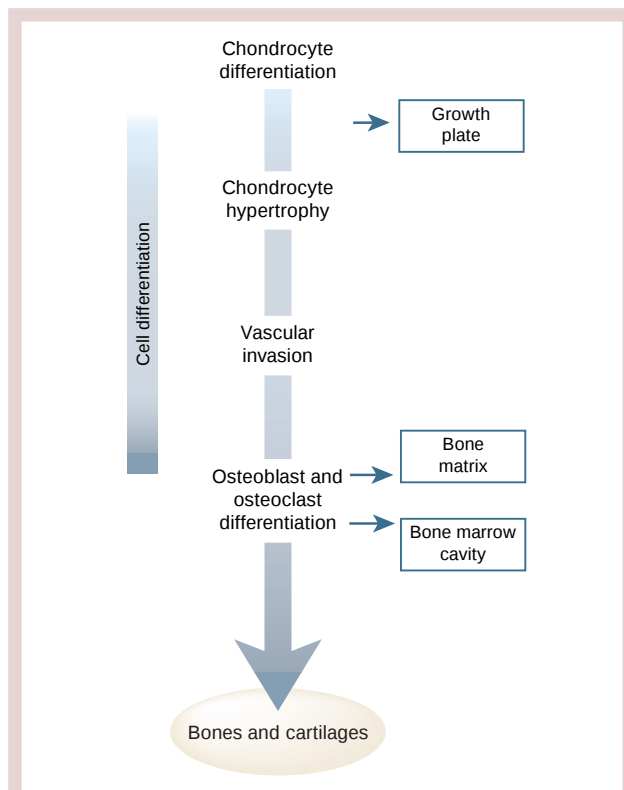


Figure 2 Each of the three specific cell types of the skeleton have particular spatial distributions. Chondrocytes are found in the growth plate and joints; subsequently they hypertrophy and die through apoptosis. Vascular invasion follows, bringing osteoblast progenitors from the bone collar into the centre of the future bone. A bone marrow then forms and osteoblasts favour osteoclast differentiation.

in a proximal-to-distal sequence⁵. A second, more recent model suggests that specification to form a given skeletal element in the developing limb occurs very early, at the time or even before the time that limb bud outgrowth is initiated⁶. This question is not fully resolved, and the outcome will undoubtedly require addressing in molecular terms another crucial yet almost unexplored question of skeletal development: what are the mechanisms leading mesenchymal cells to aggregate to form these condensations prefiguring each skeletal element at the onset of cell differentiation in the skeleton? Understanding this process will affect our understanding both of skeletal patterning and of chondrocyte differentiation.

The difficulties encountered in assembling coherent genetic pathways accounting for limb or skull patterning have been enormous and have not all been overcome. This explains why we still know so little about the genetic pathway regulating patterning of other skeletal elements such as ribs, hands and feet.

Skeletal cell differentiation

Once patterning of a given skeletal element is achieved, mesenchymal cells in this element differentiate in most cases into chondrocytes, the cell type specific of cartilage. The cartilaginous template is eventually replaced by bone containing osteoblasts and osteoclasts. This process of bone formation is called endochondral ossification. In some skeletal elements, mesenchymal cells differentiate directly into osteoblasts in a process called intramembranous ossification.

In the past ten years a combination of mouse and human genetics has provided a sophisticated understanding of how cells differentiate into chondrocyte, osteoblast (or bone-forming cells) and osteoclast (or bone-resorbing cells). Two of these lineages, the chondrocyte and the osteoblast, are of mesodermal origin and originate from a

common progenitor cell called an osteochondroprogenitor⁷. Beyond this stage, differentiation along the chondrocyte lineage will eventually give rise to resting, proliferating and hypertrophic chondrocytes that will be organized in columns in the growth plate (ref. 7, and see review in this issue by Kronenberg, page 332). Many genes encoding growth factors or transcription factors contribute to the regulation of chondrocyte differentiation. Figures 1 and 2 summarize the key issues in skeletal development. Unresolved questions concerning chondrocyte differentiation relate mainly to the molecular determinants of the columnar organization of the growth plate and how it controls longitudinal growth of the skeleton.

In contrast to what is known of chondrogenesis, only a handful of genes have so far been shown to control osteoblast differentiation. What seems to emerge from this ongoing work is that different genes control osteoblast differentiation and proliferation. For instance, *Runx2/Cbfa1* and *Osterix* control osteoblast differentiation, whereas the LRP5 (low-density lipoprotein receptor-related protein 5) signalling pathway controls osteoblast proliferation. *Runx2* is necessary and sufficient for osteoblast differentiation, although there is an unexplained delay between its expression and the appearance of bona fide osteoblasts. Currently, the only gene known that might explain this long delay is *Osterix*⁸, which seems to act downstream of *Runx2* (see reviews in this issue by Harada and Rodan, page 349, and Zelzer and Olsen, page 343). The relative paucity of transcription factors known to affect osteoblast differentiation contrasts with what is known regarding the role of these factors in other cell differentiation processes. Thus, it is likely that other transcription factors affecting osteoblast differentiation and proliferation, positively or negatively, remain to be identified. This may have important therapeutic consequences, as modulating the activity of these factors could lead to an increase in bone formation, the function that is affected in osteoporosis.

Knowledge of how osteoclast differentiation occurs is more detailed than that of osteoblast differentiation and has come from two independent lines of research that has converged over the past ten years on identical transcription factors. Gene-deletion experiments have identified a series of transcription factors — PU1, c-Fos, nuclear factor (NF)- κ B, NF- α Tc and *Mitf*^{9–13} — that control osteoclast differentiation. Another line of research identified a soluble tumour-necrosis factor receptor called osteoprotegerin as an inhibitor of osteoclast differentiation. This led rapidly to the identification of RANKL — the receptor activator of NF- κ B (RANK) ligand — as an osteoclast differentiation factor binding to the RANK receptor in osteoclasts. The RANKL signalling pathway eventually leads to NF- κ B and c-Fos activation, thus merging two aspects of osteoclast biology (see review in this issue by Boyle *et al.*, page 337). The fact that RANKL is produced by osteoblasts lends support to the concept that osteoblasts control osteoclast differentiation, not function. This, however, may be a simplistic view as RANKL is produced by many cells and is probably a circulating molecule. Nevertheless, identification of these genes has completely changed our understanding of bone resorption and may in the future affect the treatment of osteoporosis.

Skeletal physiology and pathology

Compared with our knowledge of other organs, our molecular understanding of many skeletal functions is limited. As a result, we still know little about the pathophysiology of degenerative diseases of the skeleton. One obvious difficulty in studying molecular physiology of the skeleton is that one cannot rely on conservation of regulatory pathways between *Caenorhabditis elegans*, *Drosophila* and humans, as skeleton was acquired late during evolution. Thus, answers to physiological questions have to be found using vertebrate models, mostly in mice.

The importance of studying skeletal physiology is illustrated by the incidence of degenerative diseases affecting functions such as the control and arrest of skeletal longitudinal growth, bone mineralization and the control of bone mass (Fig. 3). Osteoporosis, a disease of

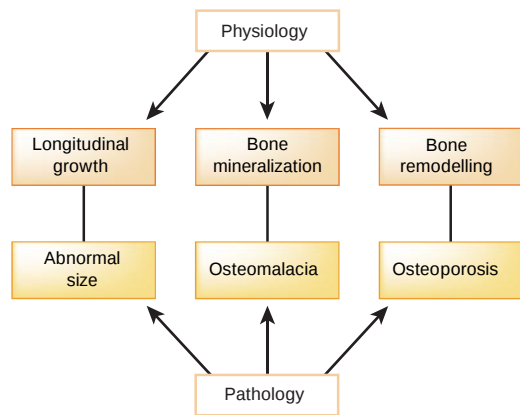


Figure 3 Unresolved issues in skeletal physiology. Three outstanding issues are: how does bone growth stop; why is mineralization restricted to bone; and how is bone mass maintained constant through bone remodelling? These impact on three types of disease in the skeleton: abnormal size, rickets and osteomalacia, and osteoporosis. In addition, understanding the molecular bases of intracellular matrix mineralization may affect our understanding of other disease such as atherosclerosis and osteoarthritis.

low bone mass, is the most frequent of all degenerative diseases and exerts a profound influence on research in skeletal biology. Furthermore, its incidence will undoubtedly increase with the ageing of the general population. At present, only symptomatic treatment is available in the form of inhibitors of bone resorption. Some of the genes affecting the control of bone mass also affect cell differentiation and proliferation during development, while other regulatory loops seem more specific to the adult skeleton. For instance, *Runx2* controls bone formation after birth by affecting osteoblast differentiation. Likewise, mutations in *LRP5*, which encodes a surface receptor, affect high bone mass in humans^{14,15}. Given the ubiquitous expression of *LRP5*, the challenge will be to identify the ligand or signal-transducing molecules that account for the skeletal action of the *LRP5* signalling pathway. Another molecule that functions during development and after birth is RANKL, which helps produce osteoclasts throughout life.

In contrast to pathways affecting skeletal differentiation, other pathways seem to affect only skeletal physiology. One of these critical regulatory loops is the negative regulation of bone resorption exerted by oestrogen, which explains the increase in bone resorption and the bone loss observed following gonadal failure at menopause. At present we do not have a clear understanding of how oestrogen regulates bone resorption and this is probably one of the most important challenges remaining in the field of skeletal physiology. It is possible that the hormone may control osteoclast function locally through a nonspecific mode of action^{16,17}.

Recent research has identified a novel means of regulation of bone mass that includes a hypothalamic relay. As discussed by Harada and Rodan on page 349, leptin is a powerful inhibitor of bone formation. Leptin acts by binding to its signal-transducing receptors present on hypothalamic neurons, with the sympathetic nervous system as its peripheral mediator. Leptin regulation of bone formation is dominant over the influence of gonads on bone resorption^{18,19}. Indeed, animals deprived of leptin signalling have non-functional gonads, and bone resorption increases as expected owing to a lack of oestrogen. But if leptin signalling is removed in this gonad-free animal, what is observed in terms of bone pathophysiology is the consequence of the absence of leptin signalling (that is, high bone mass), not the consequence of the gonadal failure. This regulation is also important for therapeutic reasons, as β -adrenergic antagonists can increase bone mass and prevent osteoporosis in

ovariectomized mice. This aspect of the regulation of bone mass has now entered the field of clinical investigation.

Lastly, how mechanical stress affects bone mass is another important aspect of skeletal physiology²⁰, and its study is taking full advantage of the many mutant mouse strains available. We have no real clues of how bones sense and transduce mechanical forces and how this may affect gene expression.

Cartilage, especially articular cartilage biology, is often ignored in reviews about skeleton biology. This is partly because we still know little about joint formation (see review in this issue by Mariani and Martin, page 319) and even less about the physiopathology of osteoarthritis, the most frequent degenerative disease of the joints. Besides osteoarthritis, autoimmune disease represents another category of joint disease, with rheumatoid arthritis being the most common example of this disease type. We have no certainty about the nature of the molecular mechanisms leading to the development of a rheumatoid arthritis (see review in this issue by Firestein, page 356). As a result, treatments are often empirical. Better understanding of the role played by cytokines involved in inflammatory processes should lead eventually to a clearer understanding of these diseases (as illustrated, for example, by studies of the role of RANKL in the development of rheumatoid arthritis). This area of skeletal biology, which lies at the frontier between bone biology and immunology, will expand rapidly as our knowledge of the molecular bases of inflammation improves.

Owing to space constraints, this Insight does not cover all aspects of skeletal biology and pathology. However, by formalizing the various questions still confronting developmental biologists, molecular physiologists and clinicians it serves to define the challenge ahead in each field. A trend that is apparent in the accompanying reviews and that should become more evident in the future is that progress in skeletal biology is having an increasing effect directly on the management of cartilage and bone diseases. □

doi:10.1038/nature01654

- Christ, B., Jacob, H. J. & Jacob, M. Experimental analysis of the origin of the wing musculature in avian embryos. *Anat. Embryol.* **150**, 171–186 (1977).
- Capdevila, J. & Izpisua Belmonte, J. C. Patterning mechanisms controlling vertebrate limb development. *Annu. Rev. Cell Dev. Biol.* **17**, 87–132 (2001).
- Tickle, C. & Munsterberg, A. Vertebrate limb development—the early stages in chick and mouse. *Curr. Opin. Genet. Dev.* **11**, 476–481 (2001).
- Niswander, L. Pattern formation: old models out on a limb. *Nature Rev. Genet.* **4**, 133–143 (2003).
- Summerbell, D., Lewis, H. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244**, 492–496 (1973).
- Dudley, A. T., Ros, M. A. & Tabin, C. J. A re-examination of proximodistal patterning during vertebrate limb development. *Nature* **481**, 539–544 (2002).
- Karsenty, G. & Wagner, E. F. Reaching a genetic and molecular understanding of skeletal development. *Dev. Cell* **2**, 389–406 (2002).
- Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108**, 17–29 (2002).
- Tondravi, M. M. *et al.* Osteopetrosis in mice lacking haematopoietic transcription factor PU.1. *Nature* **386**, 81–84 (1997).
- Grigoriadis, A. E. *et al.* c-Fos: a key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science* **266**, 443–448 (1994).
- Franzoso, G. *et al.* Requirement for NF- κ B in osteoclast and B-cell development. *Genes Dev.* **11**, 3482–3496 (1997).
- Iotsova, V. *et al.* Osteopetrosis in mice lacking NF- κ B1 and NF- κ B2. *Nature Med.* **3**, 1285–1289 (1997).
- Takayanagi, H. *et al.* T-cell mediated regulation of osteoclastogenesis by signalling cross-talk between RANKL and IFN- γ . *Nature* **408**, 600–605 (2000).
- Gong, Y. *et al.* LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell* **107**, 513–523 (2001).
- Bex, M. *et al.* The effects of growth hormone replacement therapy on bone metabolism in adult-onset growth hormone deficiency: a 2-year open randomized controlled multicenter trial. *J. Bone Miner. Res.* **17**, 1091–1094 (2002).
- Kousteni, S. *et al.* Nongenotropic, sex-nonspecific signaling through the estrogen or androgen receptors: dissociation from transcriptional activity. *Cell* **104**, 719–730 (2001).
- Kousteni, S. *et al.* Reversal of bone loss in mice by nongenotropic signaling of sex steroids. *Science* **298**, 843–846 (2002).
- Ducy, P. *et al.* Leptin inhibits bone formation through a hypothalamic relay: a central control of bone mass. *Cell* **100**, 197–207 (2000).
- Takeda, S., Bonnamy, J. P., Owne, M. J., Ducy, P. & Karsenty, G. Continuous expression of *Cbfa1* in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues *Cbfa1*-deficient mice. *Genes Dev.* **15**, 467–481 (2002).
- Turner, C. H. & Pavalko, F. M. Mechanotransduction and functional response of the skeleton to physical stress: the mechanisms and mechanics of bone adaptation. *J. Orthop. Sci.* **3**, 346–355 (1998).

Deciphering skeletal patterning: clues from the limb

Francesca V. Mariani & Gail R. Martin

Department of Anatomy and Program in Developmental Biology, School of Medicine, University of California at San Francisco, San Francisco, California 94143-0452, USA (e-mail: gmartin@itsa.ucsf.edu)

Even young children can distinguish a *Tyrannosaurus rex* from a *Brontosaurus* by observing differences in bone size, shape, number and arrangement, that is, skeletal pattern. But despite our extensive knowledge about cartilage and bone formation per se, it is still largely a mystery how skeletal pattern is established. Much of what we do know has been learned from studying limb development in chicken and mouse embryos. Based on the data from such studies, models for how limb skeletal pattern is established have been proposed and continue to be hotly debated.

Some 360 million years ago our ancestors began to crawl on land with the aid of paired appendages. Evolution has since produced limbs that permit modes of locomotion as distinct as running, flying and swinging through trees. Despite the adaptive changes to the tetrapod limb skeleton, its fundamental organization has been remarkably conserved. Three discrete compartments are evident in most extant tetrapod limbs: proximal (closest to the body wall, that is, upper arm or thigh); middle (forearm or calf); and distal (hand or foot) (Fig. 1). In the generic limb, the proximal compartment (stylopod), contains one bone, the middle (zeugopod) contains two, and the distal (autopod) contains a variable number, including at most five digits. Generally, it is variation in bone length and shape, and bone number in the autopod, that allows limbs to serve markedly different functions.

Vertebrate species can be distinguished from one another by the size, shape, number and arrangement of their bones — properties that collectively define skeletal pattern. Although the process by which skeletal pattern is established is still largely mysterious, what we do know comes primarily from studies of limb development in chicken and mouse embryos. Limb skeletal elements develop from cartilage anlagen, a process known as endochondral bone formation (see review in this issue by Kronenberg, page 332). It begins when chondrocyte progenitors form a densely packed 'condensation', which subsequently develops into a cartilage template that is ultimately replaced by bone. The essential steps in establishing limb skeletal pattern, that is, 'patterning', take place prior to and during the formation of these cartilagenous anlagen.

One of the most informative approaches for exploring the mechanism of limb patterning is to perturb limb development, and then assess the effect on skeletal pattern. Such studies have provided a wealth of intriguing data, on which models for limb skeletal patterning have been based, but much remains to be learned. Here we review what is known about limb skeletal patterning, discuss the leading models of how this process occurs and the controversy surrounding them, and identify areas that must be explored so that we can gain a deeper understanding of skeletal patterning mechanisms.

Key steps in limb skeletal patterning

Each limb arises from a small bud of mesodermal cells covered by surface ectoderm. The mesodermal cell population

comprises all the progenitors of the chondrocytes and connective tissues such as tendons and muscle sheaths. Other tissues, such as muscle and blood vessels, develop from cells that migrate into the early limb bud^{1–4}.

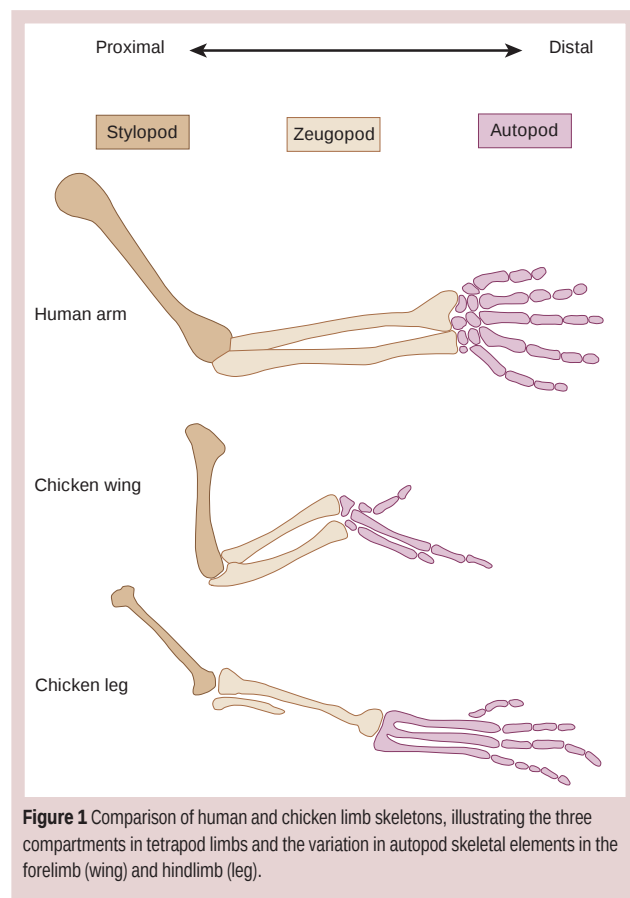
Establishment and outgrowth of the limb bud

One potentially important factor in defining skeletal pattern is the number of cells that constitute the initial limb bud, but little is known about how this number is determined. In part, initial limb bud size is controlled by the signalling system that induces limb bud formation at stereotypic positions along the body axis (Fig. 2a). Such signals, which originate in tissues that lie interior (medial) to the lateral plate mesoderm (LPM) from which the limb bud forms, regulate gene expression in the LPM^{5–7}. However, a signal from the ectoderm that covers the limb-forming LPM also seems to be involved. Initial limb bud size is significantly reduced when FGF8, a member of the fibroblast growth factor (FGF) family of secreted signalling molecules, is inactivated in prospective limb ectoderm⁸. The limbs that develop in such *Fgf8*-deficient mice lack several skeletal elements^{9,10}, defects that could be due to small initial bud size and/or to lack of FGF8 signalling at later stages.

Once established, the limb bud rapidly increases in size (Fig. 2b). Fate mapping studies have shown that expansion of the progenitors of the proximal, middle and distal compartments is completed at different times, in a proximal to distal sequence^{11,12}. Little is known about how this is regulated, but it may be a reflection of the timing of condensation, which is correlated with decreased mitotic activity¹³.

Condensation

The formation of the condensations that prefigure the skeletal elements is arguably the most critical event in skeletal patterning, so it is surprising how little is known about what determines their size, shape and number. Prechondrogenic condensations are detectable morphologically as local regions of increased cell density (Fig. 2c), and by molecular markers¹⁴. Cell–cell interactions are important in initiating condensation, perhaps by establishing an 'aggregation centre' that recruits cells from surrounding tissue¹⁵. At the molecular level, evidence largely from cell-culture studies suggests that several different classes of molecules, including the cell adhesion molecules N-cadherin and NCAM, are important in condensation^{14,16}.



It appears that condensations form at specific stages of limb development, in a proximal to distal sequence, irrespective of how many chondrocyte progenitors are present. If the timing of condensation is indeed fixed, then alterations in cell number in the limb bud should affect skeletal pattern. For example, if fewer chondrocyte progenitors are present at the time of condensation, one might expect smaller elements to form. In fact, there is evidence suggesting that when limb bud cell number is reduced, the skeletal elements that form are smaller than normal and misshapen, and when it drops below a certain threshold, condensations do not develop^{17,18}. If more chondrocyte progenitors are present at the time of condensation, one might expect larger elements to form. Interestingly, however, mutants with limb buds that are larger than normal, and thus presumably contain more cells than normal, develop limbs with more elements in the autopod¹⁹.

A few genes required for condensation have been identified by mutational analysis in mice. For example, null mutations in *Bmp5*, a member of the bone morphogenetic protein (BMP) family of signalling molecules, result in the absence or alteration in the shape of specific condensations at various locations in the body (but not in the limbs)²⁰. More global effects, leading to the absence of all condensations, are observed when *Sox9*, an Sry-related transcription factor gene, is inactivated in the early mouse limb bud²¹. However, it is not yet known what role *Sox9* plays in condensation.

One of the many unresolved questions about the mechanism of condensation in the limb bud is whether multiple focal condensations form at different locations, or if an initial focal condensation forms and then elongates at its distal end, branching at various points to increase the number of skeletal element precursors across the width of the limb (for example, at the stylopod/zeugopod boundary, where there is a transition from one to two elements; see Fig. 1)¹⁵. The argument against the latter model is that there are several experimen-

tal situations in which distal structures form, apparently in the absence of any proximal structures from which they could be derived by branching²².

Joint formation

Joint formation is another process crucial for patterning. It occurs in several stages²³, either at the boundary between two adjacent condensations or within a single condensation. For example, in the autopod, individual digital rays are divided by joints into hand or foot elements and the phalanges that comprise the digits (Fig. 2d). Therefore, proper positioning of joints within a condensation influences the number and size of skeletal elements.

Joints become morphologically detectable when chondrocytes in the prospective joint-forming region become denser and flatten, and chondrogenic differentiation is inhibited. This creates an 'interzone', the site of the future cell death that creates the joint space. Two genes, *Gdf5* and *Nog*, which encode a BMP-related protein and BMP antagonist, respectively, are crucial in joint formation. Loss of *Gdf5* function results in several skeletal abnormalities including absence of specific joints in the autopod²⁴. Further analysis has suggested that *Gdf5* has multiple roles in skeletal development, including restriction of joint development to the appropriate locations²⁵. Loss of *Nog* function causes complete failure of joint formation in the autopod, at least in part via an effect on *Gdf5* expression²⁶. Ectopic expression of *Wnt14* in chicken limb buds induces morphological and molecular manifestations of early joint formation, including expression of *Gdf5*, suggesting that this member of the Wntless family of intercellular signalling molecules also has a role in normal joint formation²⁷. These and related studies have provided a good start towards understanding the molecular mechanism of joint formation and how they are localized to specific sites.

Cartilage and bone formation

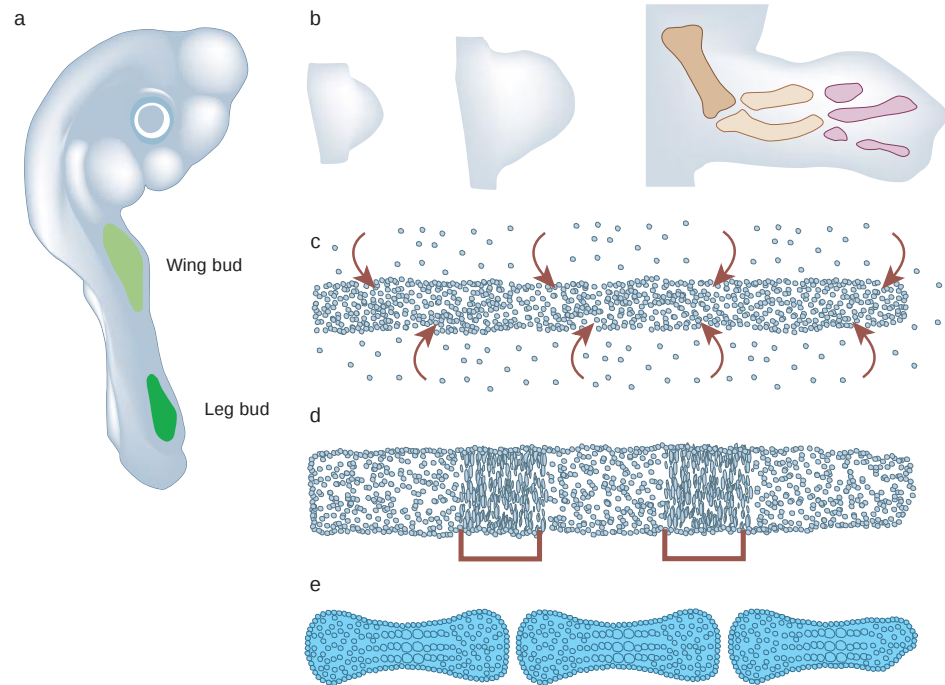
Much of the patterning process is complete by the time that condensations have differentiated into cartilaginous templates (Fig. 2e) and cartilage markers such as Alcian blue are detectable. However, the remaining steps in endochondral bone formation do influence final skeletal pattern, and patterning defects are observed as a consequence of experimental manipulations or mutations that perturb these late-stage processes. For example, mutations in *Ror2*, a receptor-like tyrosine kinase, cause limb skeletal defects as a result of abnormal cartilage development²⁸.

Regulatory genes affecting pattern

A process as intricate as limb skeletal patterning obviously depends on the deployment of numerous regulatory genes, including homeobox transcription factor (Hox) genes related to the *Drosophila* homeotic genes. Vertebrate Hox complexes arose by two sequential duplications of an ancestral cluster, and therefore genes at corresponding positions in the different complexes are similar in sequence (paralogues)²⁹. Their distinctive expression patterns in the developing limb prompted investigators to undertake mutational analysis in mice, which has shown that genes in paralogue groups 9–13 of the *Hoxa* and *Hoxd* clusters are crucial in patterning the limb skeleton. At present, the view is that *Hoxa* and *Hoxd* group 9 and 10 genes influence stylopod size and shape, group 11 genes are essential for normal development of zeugopod elements, and group 11, 12 and 13 genes are involved in patterning the autopod, which is extremely sensitive to the level of expression of these genes³⁰. But interpretation of the data is complicated by the fact that *Hoxd*-null mutant phenotypes vary markedly depending on whether the allele was produced by insertion or deletion. These differences can be explained by the finding that gene deletion, but not insertional mutation, causes alterations in the expression of nearby genes that have profound effects on skeletal element number, size and shape³¹.

Because *Hoxa* and *Hoxd* genes are expressed in dynamic patterns throughout limb development³², an important question is when do

Figure 2 Key stages in limb skeletal development. **a**, Limb formation initiates in the chicken embryo at ~3 days of development, when the primary embryonic axis (head to tail) is elongating. First the forelimb (wing) buds and then the hindlimb (leg) buds begin to protrude from the sides of the embryo. **b**, Substantial outgrowth and patterning occurs over the course of the next three to four days, culminating in the establishment of the cartilage templates that prefigure the bones (illustrated on the right). **c**, During this time, chondrocyte progenitors aggregate (brown arrows) and form prechondrogenic condensations (a schematic digital ray is illustrated); **d**, joint formation begins (brown brackets); and **e**, the chondrocytes differentiate and begin secreting cartilage extracellular matrix molecules, which can be detected by Alcian blue staining.



they act to affect limb pattern? Several studies have pointed to effects on condensation or later events. For example, misexpression of *Hoxd13* in chicken limb buds leads to a decrease in cell proliferation in the growth plate of the tibia³³. Moreover, in mice lacking both *Hoxa11* and *Hoxd11* function, the cartilage templates that will form zeugopod elements are nearly normal in size at early stages of their development, but in the final skeleton the zeugopod is dramatically reduced³⁴.

Other transcription factor genes, including members of the T-box family, also are important in limb skeletal patterning^{35–39}. In fact, based on overexpression studies in chicken limb buds, it has been suggested that forelimb and hindlimb identity are specified by *Tbx5* and *Tbx4*, respectively⁵. However, analysis of *Tbx5*- and *Tbx4*-null mutant mouse embryos has not provided any insight into this issue, because forelimbs do not develop in the absence of *Tbx5* function⁴⁰, and without *Tbx4* function, although limb buds form, embryos do not survive long enough to determine limb skeletal pattern (V. Papaioannou, personal communication).

Sources of signals affecting pattern

Over the past half-century, experimental manipulations, primarily on chicken limb buds, have provided evidence that specific regions of the limb bud produce signalling molecules that influence limb skeletal pattern^{41–43}. The knowledge that such regions exist in the early limb bud has led to the hypothesis that key steps in skeletal patterning occur as limb bud outgrowth proceeds. Here we highlight basic information about these signalling regions; more comprehensive treatments can be found in recent reviews^{5,44}.

Apical ectodermal ridge

The apical ectodermal ridge (AER) is a morphologically distinct ectoderm that rims the distal tip of the limb bud (Fig. 3a). Interest in AER function began in 1948, when Saunders⁴⁵ removed the AER from the chicken limb bud at successive stages of development and found that this caused abnormalities in skeletal pattern. The experimental limbs looked as though they had been severed (truncated) at different distances from the shoulder. When the AER was removed

early, zeugopod and autopod were absent, but when it was removed later, only autopod was absent (Fig. 3b), showing that signals from the AER are essential for skeletal development. Given the importance of the AER, the mechanism by which it is established and maintained has been the subject of intense study^{5,6,42,46}.

Numerous genes expressed in the AER have been identified, including several members of the FGF and BMP families^{5,6,42} (Fig. 3c). Application of beads containing recombinant FGF protein to the distal tip of AER-excised limb buds rescues skeletal development^{47,48}, and no other molecules have been reported to have such activity. Thus, FGFs are considered to be the key mediators of AER function.

Recent studies have provided an explanation for the phenotypes caused by AER removal and the ability of FGFs to functionally compensate for it. Following up on earlier experiments showing that AER removal causes cell death^{47,49}, Dudley *et al.*¹² found that the domain of cell death remains constant in size (extending inwards ~200 μ m from the distal tip) when the AER was removed at early and mid limb bud stages. When it was removed at much later stages, distal cells survived but had a significantly lower than normal mitotic index. This suggests that the truncations are progressively less severe when the AER is removed at later and later stages, because over time, the domain of cells affected contains a smaller and smaller proportion of distal limb skeletal progenitors. It was also found that FGF protein rescued skeletal development only when applied before cells died, and moreover when cells were dye-labelled in the region where cell death was observed, they failed to contribute to skeletal elements. These data support the hypothesis that AER removal causes distal truncations because it results in death of the distal cells at early stages or reduces their proliferation rate at late stages, and that FGF rescues limb skeletal development by preventing these effects.

To determine the function of FGFs produced specifically in the AER (AER-FGFs) during normal limb development, gene knock-out experiments have been performed in mice. When two of the four AER-FGFs, FGF4 and FGF8, are eliminated in the limb bud using a conditional gene inactivation approach^{8–10,50,51}, no limb forms. However, when AER-FGF function is present early and then eliminated, limbs develop with abnormal skeletal pattern, and the defects,

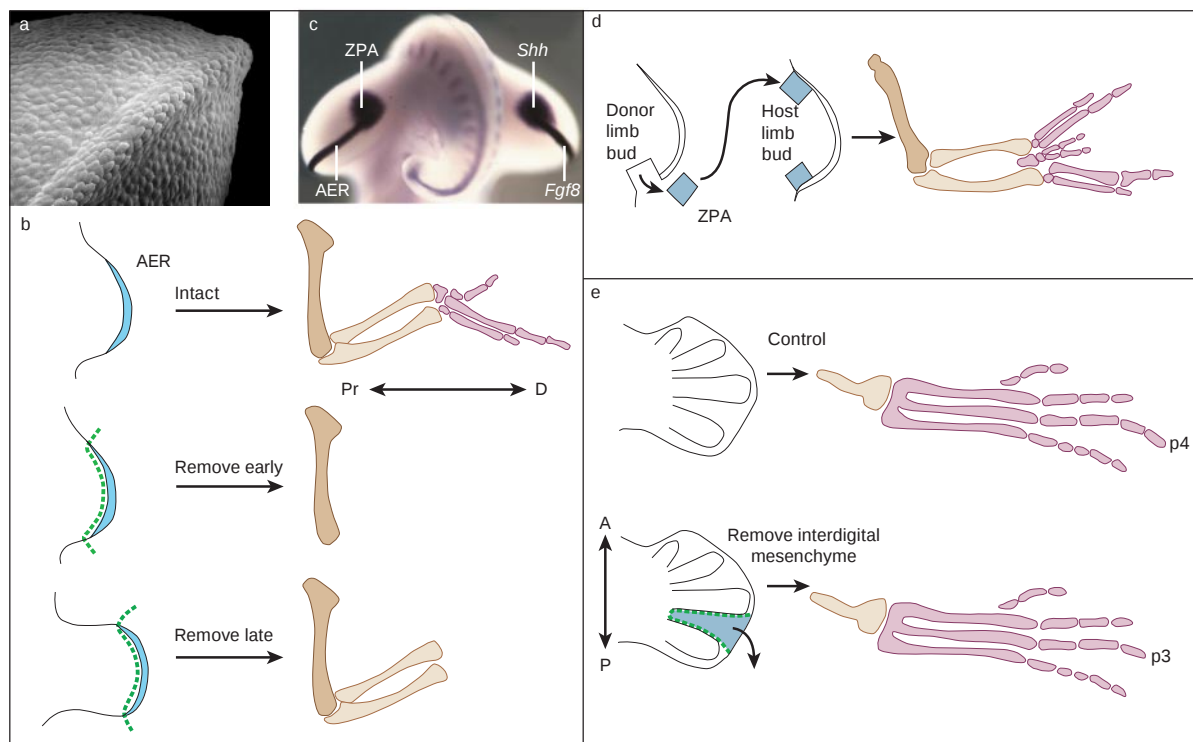


Figure 3 Signalling centres in the limb bud. **a**, Scanning electron micrograph showing the distinctive morphology of the apical ectodermal ridge (AER) that is localized at the distal tip of the limb bud (photograph courtesy of K. W. Tosney). **b**, Removal of the chick wing bud AER at an early stage results in limbs that appear severed (truncated) at the level of the elbow. AER removal at a later stage results in truncation at a more distal level⁴⁵. **c**, Expression of genes encoding secreted signalling molecules, as visualized by RNA *in situ* hybridization in this posterior view of the leg buds of a chicken embryo. *Fgf8* is expressed in the AER and *Shh* in the zone of polarizing activity (ZPA). The tail is

marked by *Shh* expressing cells along its length. (Photograph courtesy of P. Crossley.) **d**, When the ZPA is removed from its normal location on the posterior side of a donor limb bud and grafted to the anterior side of a host limb bud, the resulting limb has a mirror image duplication of the autopod skeletal elements⁵² (compare with normal pattern of wing skeletal elements in panel **b**). **e**, Removal of the interdigital mesenchyme at early post-condensation stages alters the number of phalanges in the digit anterior to the deleted region⁶². In this case a digit that normally has four phalanges (p4 in control) now has only three.

including a decrease in skeletal element number or size, are traceable to pre- or early condensation stages⁸. Based on a detailed analysis of the mutant phenotypes, AER-FGF signalling seems to influence skeletal pattern, at least in part, by regulating the number of chondrocyte progenitors available for condensation via an effect on cell survival.

Zone of polarizing activity

The zone of polarizing activity (ZPA) resides in mesenchyme at the posterior distal margin of the limb bud⁵². Grafting a ZPA to the anterior distal margin of a chicken limb bud causes mirror-image duplications of skeletal elements (Fig. 3d), suggesting that this region controls patterning along the anterior–posterior axis (thumb to little finger)⁵³. The secreted protein Sonic hedgehog (SHH) is produced in the ZPA (Fig. 3c), and *Shh* expression can mimic the effects of ZPA grafts, suggesting that it can be equated with ZPA activity⁵⁴. It is thought that SHH functions as a morphogen, possibly through an inductive effect on *Bmp2* expression⁵⁵, and that the distance it diffuses depends upon post-translational modifications⁵⁶. The molecular mechanism by which SHH signalling regulates its target genes involves effects on GLI3, the vertebrate homologue of *Drosophila* Cubitus interruptus, which functions either as transcriptional activator (GLI3A) or repressor (GLI3R). High levels of SHH prevent the formation of GLI3R and promote GLI3A function. Thus the local concentration of SHH regulates target gene expression by controlling the balance of GLI3 repressor and activator forms^{57,58}.

Mutational analysis in mice has provided clues about the normal function of SHH in limb skeletal development. *Shh*^{−/−} mutant limbs

have a normal stylopod, but exhibit severe defects distally^{59,60}. The presence of an extensive domain of abnormal cell death in *Shh*^{−/−} limb buds⁵⁹ suggests that, like AER-FGFs, SHH may influence skeletal pattern by controlling cell survival. The ectopic apoptosis in *Shh*^{−/−} limb buds is caused by the abnormally high level of GLI3R produced in the absence of SHH, because removing one functional copy of *Gli3* in *Shh*^{−/−} mice reduces the amount of ectopic apoptosis, and removing both functional copies of *Gli3* in *Shh*^{−/−} mice eliminates it⁶¹. When *Gli3* dosage is reduced in *Shh*^{−/−} mice, zeugopod and autopod skeletal elements develop, presumably because there are now sufficient chondrocyte progenitors available to form the condensations. As expected, *Shh*^{−/−}*Gli3*^{+/−} mice have fewer digits (3–4) than *Shh*^{−/−}*Gli3*^{−/−} mice (6–11 digits)^{57,61}. The digits formed in *Shh*^{−/−} limbs in the absence of one functional copy of *Gli3* all resemble digit 1 (as judged by phalanx number), whereas those that form in the absence of both *Shh* and *Gli3* function are morphologically indistinguishable from one another, and difficult to categorize. These observations show that in addition to regulating cell survival, SHH also has a role in patterning the autopod.

Interdigital mesenchyme

Although the potential patterning function of signals produced at early stages has been the subject of numerous studies, little is known about how signals produced at later stages, after initiation of condensation, affect skeletal pattern. Dahn and Fallon⁶² demonstrated that the mesenchyme between the digital rays (interdigital mesenchyme or IDM) produces signals that influence the number of phalanges

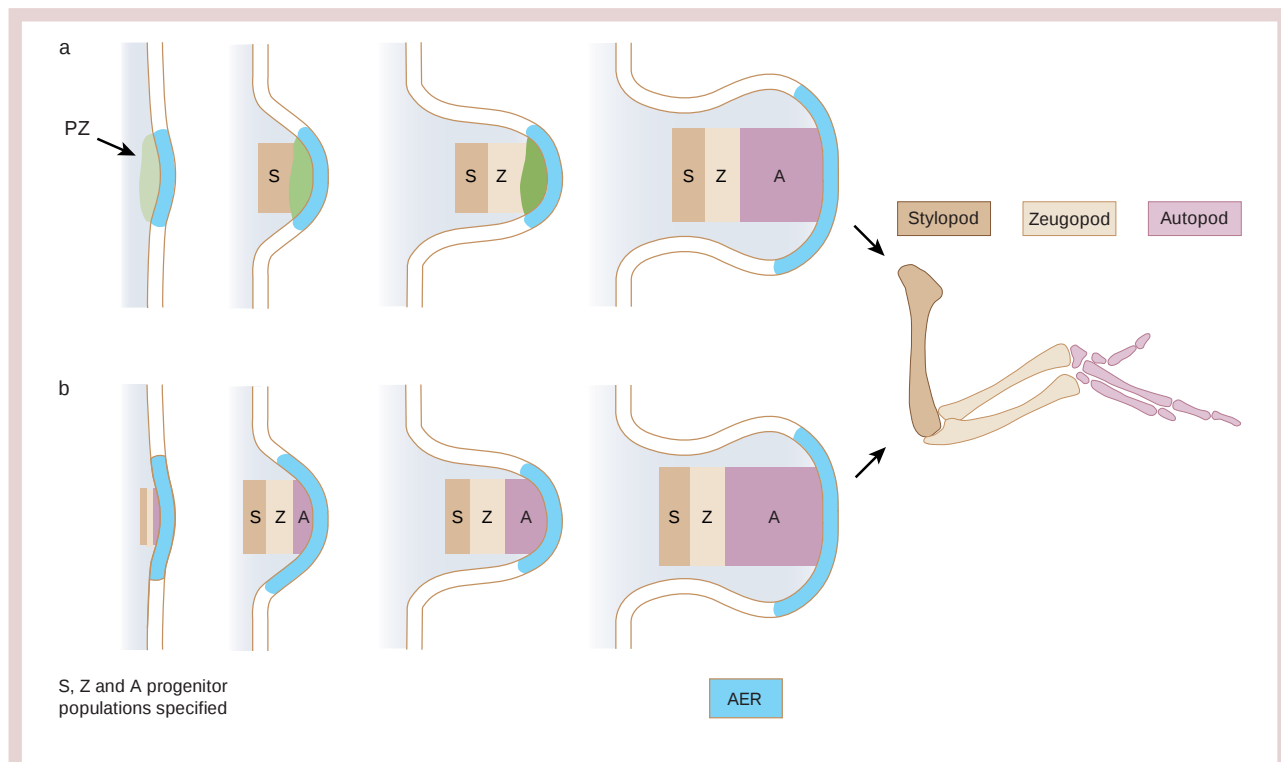


Figure 4 Two models for proximodistal limb skeletal patterning. **a**, The progress zone model proposes that the mesoderm at the distal tip of the limb bud constitutes a 'progress zone' (PZ, in green) of fixed dimensions, in which cells are receiving progressively more distal positional information over time (as indicated by the change in intensity of green colour from one stage to the next). When cells exit the progress zone they lose their lability, as they are no longer under the influence of signals from

the AER (blue). Thus, the first to leave become stylopod (S) progenitors, whereas those that leave later become zeugopod (Z) and even later, autopod (A) progenitors⁶⁵. **b**, The early specification model¹² proposes that at an early limb bud stage, cells are broadly specified to form the three compartments of the limb. The cells then undergo considerable expansion before becoming determined to form the different skeletal elements, as indicated.

that ultimately develop (Fig. 3e). These signals apparently include BMPs, because application of Noggin protein causes alterations in phalanx number similar to those observed when the IDM is removed. These effects on digit pattern underscore the importance of local signalling during cartilage template development, and raise the possibility that signals from cells surrounding more proximal structures may also influence skeletal pattern. Such local sources of signalling molecules at late stages may be established as a consequence of signalling at early stages by SHH or other factors.

Models of limb skeletal patterning

Efforts to explain the mechanism of limb patterning have been based on two fundamental concepts in developmental biology — specification and determination⁶³. A cell or group of cells is defined as 'specified' when it has acquired molecular information that allows it to differentiate into a particular cell or tissue type, and, when placed in 'neutral' environments, will differentiate into that particular cell or tissue type. Specified cells are still labile, because they can respond to new signals by altering their differentiation path, that is, they can be 're-specified' to form a different cell or tissue type. As development proceeds, cells lose this lability, and finally become 'determined'. Determined cells differentiate into the specified cell or tissue type, even when they are exposed to new signals.

The prevailing view of limb patterning is that at early stages cells become specified to form different parts of the limb skeleton. Although there are obvious mechanisms for specifying a particular cell or tissue type, the mechanism for specifying structure is less clear. To address this problem, it has been proposed that cells acquire 'positional information', which tells them where they are with respect to certain boundary or reference regions. 'Positional values' in a

three-dimensional coordinate system acquired at early stages are later interpreted by appropriate cytodifferentiation⁶⁴. Specification to form a particular cell type and acquisition of positional information may occur independently. Furthermore, anterior–posterior and proximodistal patterning mechanisms are thought to be separable, but obviously need to be coordinated in some way.

The 'progress zone' model proposed by Wolpert and colleagues^{65,66} postulates that cells acquire proximodistal positional information progressively, in a proximal-to-distal sequence, by measuring time spent in a 'progress zone' (PZ) ~300 µm deep from the limb bud tip. The longer the time spent in the PZ, the more distal the positional values they acquire. As all cells in the PZ are dividing, but PZ size is constant, cells must be continually exiting the PZ, at which time the specification clock stops. Cells that exit early are specified to form proximal limb structures, whereas those that exit late are specified to form distal ones (Fig. 4a). According to this model, the AER facilitates specification by maintaining PZ cells in a labile state. Because FGFs can functionally replace the AER^{47,48}, they are assumed to be the factors that maintain such lability. However, some of the phenotypes observed when AER-FGFs are inactivated do not match what would be expected if the proximodistal specification clock were slowed or stopped⁸.

Early efforts to test the PZ model involved grafting the distal tip of an 'old' limb bud to the proximal stump of a 'young' limb bud. This model predicts that the composite bud should develop into a limb missing middle structures (zeugopod), because old distal tip cells should have already distalized beyond the point of producing zeugopod, and there would be no stimulus for cells in the stump to form a zeugopod. Although this result was obtained by one group of investigators⁶⁵, the opposite result — a complete proximodistal axis

including zeugopod elements — was obtained by others^{67,68}. The reason for these different outcomes is not known.

One popular misconception about the PZ model is that cells in the hypothetical PZ drive outward growth of the limb bud by proliferating more rapidly than cells elsewhere in the limb bud. But even before the PZ was proposed, Hornbruch and Wolpert⁶⁹ showed that the mitotic index is relatively uniform throughout early chicken limb buds, and the same has been found in mice⁸. Another misconception is that the PZ model is a stem cell model, which postulates that when individual PZ cells divide, they give rise to one daughter cell that exits the PZ proximally and one that remains within it. If so, descendants of individual PZ cells should contribute to skeletal elements along the length of the proximodistal axis. However, the PZ model was never envisaged by its authors as a stem cell model; they hypothesized that at any stage, cells localized in the deepest part of the PZ soon find themselves outside the PZ and therefore cease changing positional values. Consistent with this view, fate-mapping studies show that descendants of small groups of cells at different depths within the proposed PZ contribute almost exclusively to one skeletal compartment^{11,12}.

This last observation prompted Dudley *et al.*¹² to propose that specification to form proximal, middle or distal limb structures does not occur progressively, but instead occurs very early, perhaps even before limb bud outgrowth has begun. According to this 'early specification model' (Fig. 4b), the specified populations subsequently expand as the limb bud grows, and become determined in a proximal-to-distal sequence. This concept has been criticized on the grounds that there are not enough cells in the early limb bud to comprise a miniature limb skeleton that will simply expand to full size over time⁷⁰. But such criticism is based on the assumption that the early specification model proposes that all positional values are assigned early; in contrast, Dudley *et al.*¹² suggest that cells are specified more broadly, as progenitors of stylopod, zeugopod or autopod, and that over time they expand and become determined to form particular structures in response to cell–cell interactions and signalling.

The PZ and early specification models have different explanations for the results of various studies (see Box 1), but debate continues as to whether one or the other — or neither — model is correct. This issue could be resolved if molecular markers were found that could distinguish cells specified to form one set of limb skeletal elements from another. Such markers could then be used to determine the timing of specification.

If one accepts the premise that there is a mechanism for specifying cells in the early limb bud to form particular skeletal elements, then the key question — which is almost never discussed — is what is the link between the patterning information acquired at early stages and the morphogenetic processes, such as condensation and joint formation, that realize skeletal pattern? One possibility is that patterning information influences the timing of condensation. The earlier it occurs, the smaller the number of cells that will be available for condensation. Conversely, the later it occurs, the greater the number of cells that will be available. And, as discussed above, this could profoundly affect skeletal pattern by influencing the final size and number of elements that form. It is also possible that patterning information acquired at early limb bud stages influences condensation location. In considering the various possibilities, it is important to bear in mind that it is not necessarily the chondrocyte progenitors, but other mesodermal cell types in the limb bud that acquire patterning information. In turn, they might produce signals that influence condensation, cartilage development or joint formation, and therefore skeletal pattern.

A radically different view is that cells in the early limb bud, although perhaps specified to form chondrocytes, have no information about what particular skeletal elements they or their descendants will ultimately form. In this view, the events that occur during early limb development are permissive rather than instructive, producing a bud of appropriate size and shape, within which patterning is initiated as condensations begin to form. Skeletal

Box 1

Progress zone versus early specification models

For 30 years the 'progress zone' (PZ) model (Fig. 4a) has been the prevailing hypothesis to account for proximodistal limb patterning. According to this model, the apical ectodermal ridge (AER) is a source of permissive signals that keep cells in the PZ labile so they can autonomously acquire progressively more distal positional information. Recently, Dudley *et al.*¹² have proposed an alternative model, in which specification to form the different compartments of the limb is not progressive, but instead occurs at a very early stage of limb bud development (Fig. 4b). The two models propose different explanations for the results of various experimental manipulations, as indicated in the table below.

Box 1 Table Explanations of experimental manipulations

PZ model ^{70,73}	Early specification model ¹²
AER removal at successive stages causes distal truncations at increasingly more distal positions along the length of the limb ⁴⁵ (see Fig. 3b)	
In the absence of the AER, cells lose their lability and therefore can no longer change positional values. The later the AER is removed, the more distal the structures that have been specified before the distalizing clock stops	AER removal causes rapid cell death or a significant decrease in cell proliferation in a domain of fixed size (around the size of the hypothetical PZ). Over time, the domain of cells affected contains a progressively smaller proportion of distal limb skeletal progenitors. (It is noteworthy that unlike all the other explanations listed in this Table — on both sides of the argument — this is the only one that is based on experimental observations rather than hypothesis.)
In limb buds exposed to X-irradiation, proximal skeletal elements are severely reduced, whereas distal ones are less affected ¹⁸ .	
To compensate for X-ray-induced cell death, the surviving cells must spend extra time in the PZ, while they proliferate and restore the PZ to normal size. They therefore undergo more distalization than normal, and consequently fail to form proximal elements.	Because the specified cells differentiate in a proximal-to-distal sequence at specific developmental stages, proximal cells have less time to replenish the population in the interim between X-irradiation and the onset of condensation than do distal cells. Thus proximal cells are less likely than distal cells to form normal elements.
When a sliver of the tip of a young limb bud (including the AER) is grafted to another location, it gives rise only to digit-like skeletal elements ¹² .	
To compensate for the small number of cells in the graft, the cells proliferate, stay longer in the PZ, and thus become specified to form only digits.	Cells at the distal tip of the early limb bud are already specified to form autopod elements and realize their developmental potential in the graft.
When distalmost cells from several limb buds are dissociated, reaggregated, placed in a limb bud ectodermal jacket, and grafted to a host embryo, different outcomes are obtained depending on the developmental age of the donor limb buds: stage 20 distal cells formed stylopod, zeugopod and autopod elements; stage 22 distal cells formed zeugopod and autopod elements; stage 24 distal cells formed only autopod elements ¹² .	
A new PZ is established in the reaggregate limb buds. The different outcomes are due to the different specification states of cells taken at progressively later stages.	Although already specified, distal cells taken at early stages can be respecified in the reaggregates, but become progressively refractory to respecification.

element size and shape are roughly established as a function of basic cellular processes such as aggregate formation by chondrocyte progenitors⁷¹ and the dimensions of the domains from which cells are recruited to aggregates. Their final form is then determined by local cell–cell interactions and signalling^{62,72}.

Perspectives and future directions

However valuable the models proposed to date might be, none of them provides more than a rudimentary sketch of how skeletal pattern is achieved. Although progress will doubtless be made by continuing to study the early limb bud, equal emphasis needs to be

placed on acquiring a deeper understanding of cellular and molecular events at later stages. For example, using very early markers of chondrocyte identity and new techniques for studying cell behaviour *in vivo*, in both normal and mutant embryos, it should be possible to answer vital questions such as what controls the timing of condensation, and the dimensions and locations of the condensations. Likewise, it will be important to explore how condensations are transformed into cartilage templates, and joints are positioned.

Using sophisticated genetic approaches that allow for gene inactivation at different developmental stages, it should be possible to determine whether genes that regulate skeletal pattern, such as the *Hoxa* and *Hoxd* genes, are required at early limb bud stages, or if they act only at later stages when skeletal pattern is being realized. With so many avenues to explore, we should soon gain more extensive knowledge of how limb skeletal pattern is established in model organisms. In turn, such information should further our understanding of how variation in skeletal pattern among different species is achieved. □

doi:10.1038/nature01655

- Christ, B., Jacob, H. J. & Jacob, M. Experimental analysis of the origin of the wing musculature in avian embryos. *Anat. Embryol.* **150**, 171–186 (1977).
- Chevallier, A., Kieny, M. & Mauger, A. Limb-somite relationship: origin of the limb musculature. *J. Embryol. Exp. Morphol.* **41**, 245–258 (1977).
- Wilting, J. *et al.* Angiogenic potential of the avian somite. *Dev. Dyn.* **202**, 165–171 (1995).
- Kardon, G., Campbell, J. K. & Tabin, C. J. Local extrinsic signals determine muscle and endothelial cell fate and patterning in the vertebrate limb. *Dev. Cell* **3**, 533–545 (2002).
- Capdevila, J. & Izpisua Belmonte, J. C. Patterning mechanisms controlling vertebrate limb development. *Annu. Rev. Cell Dev. Biol.* **17**, 87–132 (2001).
- Tickle, C. & Munsterberg, A. Vertebrate limb development—the early stages in chick and mouse. *Curr. Opin. Genet. Dev.* **11**, 476–481 (2001).
- Martin, G. Making a vertebrate limb: new players enter from the wings. *BioEssays* **23**, 865–868 (2001).
- Sun, X., Mariani, F. V. & Martin, G. R. Functions of FGF signalling from the apical ectodermal ridge in limb development. *Nature* **418**, 501–508 (2002).
- Lewandoski, M., Sun, X. & Martin, G. R. Fgf8 signalling from the AER is essential for normal limb development. *Nature Genet.* **26**, 460–463 (2000).
- Moon, A. M. & Capecchi, M. R. Fgf8 is required for outgrowth and patterning of the limbs. *Nature Genet.* **26**, 455–459 (2000).
- Vargesson, N. *et al.* Cell fate in the chick limb bud and relationship to gene expression. *Development* **124**, 1909–1918 (1997).
- Dudley, A. T., Ros, M. A. & Tabin, C. J. A re-examination of proximodistal patterning during vertebrate limb development. *Nature* **418**, 539–544 (2002).
- Janners, M. Y. & Searls, R. L. Changes in rate of cellular proliferation during the differentiation of cartilage and muscle in the mesenchyme of the embryonic chick wing. *Dev. Biol.* **23**, 136–165 (1970).
- Hall, B. K. & Miyake, T. All for one and one for all: condensations and the initiation of skeletal development. *BioEssays* **22**, 138–147 (2000).
- Oster, G., Shubin, N., Murray, J. & Alberch, P. Evolution and morphogenetic rules: the shape of the vertebrate limb in ontogeny and phylogeny. *Evolution* **42**, 862–884 (1988).
- DeLise, A. M., Fischer, L. & Tuan, R. S. Cellular interactions and signaling in cartilage development. *Osteoarthritis Cartilage* **8**, 309–334 (2000).
- Gruneberg, H. *The Pathology of Development. A study of Inherited Skeletal Disorders in Animals* (Blackwell Scientific Publications, Oxford, 1963).
- Wolpert, L., Tickle, C. & Sampford, M. The effect of cell killing by X-irradiation on pattern formation in the chick limb. *J. Embryol. Exp. Morphol.* **50**, 175–193 (1979).
- Dvorak, L. & Fallon, J. F. *Talp2* mutant chick limb has anteroposterior polarity and altered patterns of programmed cell death. *Anat. Rec.* **231**, 251–260 (1991).
- King, J. A., Marker, P. C., Seung, K. J. & Kingsley, D. M. BMP5 and the molecular, skeletal, and soft-tissue alterations in *short ear* mice. *Dev. Biol.* **166**, 112–122 (1994).
- Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & de Crombrughe, B. The transcription factor *Sox9* has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of *Sox5* and *Sox6*. *Genes Dev.* **16**, 2813–2828 (2002).
- Cohn, M. J., Lovejoy, C. O., Wolpert, L. & Coates, M. I. Branching, segmentation and the metapterygial axis: pattern versus process in the vertebrate limb. *BioEssays* **24**, 460–465 (2002).
- Francis-West, P. H. *et al.* Mechanisms of GDF-5 action during skeletal development. *Development* **126**, 1305–1315 (1999).
- Storm, E. E. *et al.* Limb alterations in brachypodism mice due to mutations in a new member of the TGF- β superfamily. *Nature* **368**, 639–643 (1994).
- Storm, E. E. & Kingsley, D. M. GDF5 coordinates bone and joint formation during digit development. *Dev. Biol.* **209**, 11–27 (1999).
- Brunet, L. J., McMahon, J. A., McMahon, A. P. & Harland, R. M. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science* **280**, 1455–1457 (1998).
- Hartmann, C. & Tabin, C. J. Wnt-14 plays a pivotal role in inducing synovial joint formation in the developing appendicular skeleton. *Cell* **104**, 341–351 (2001).
- DeChiara, T. M. *et al.* *Ror2*, encoding a receptor-like tyrosine kinase, is required for cartilage and growth plate development. *Nature Genet.* **24**, 271–274 (2000).
- Krumlauf, R. Hox genes in vertebrate development. *Cell* **78**, 191–201 (1994).
- Zakany, J. & Duboule, D. Hox genes in digit development and evolution. *Cell Tissue Res.* **296**, 19–25 (1999).
- Kmita, M., Fraudeau, N., Herault, Y. & Duboule, D. Serial deletions and duplications suggest a mechanism for the collinearity of *Hoxd* genes in limbs. *Nature* **420**, 145–150 (2002).
- Nelson, C. E. *et al.* Analysis of *Hox* gene expression in the chick limb bud. *Development* **122**, 1449–1466 (1996).
- Goff, D. J. & Tabin, C. J. Analysis of *Hoxd-13* and *Hoxd-11* misexpression in chick limb buds reveals that *Hox* genes affect both bone condensation and growth. *Development* **124**, 627–636 (1997).
- Davis, A. P., Witte, D. P., Hsieh-Li, H. M., Potter, S. S. & Capecchi, M. R. Absence of radius and ulna in mice lacking *hoxa-11* and *hoxd-11*. *Nature* **375**, 791–795 (1995).
- Bamshad, M. *et al.* Mutations in human *TBX3* alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nature Genet.* **16**, 311–315 (1997).
- Bruneau, B. G. *et al.* A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor *Tbx5* in cardiogenesis and disease. *Cell* **106**, 709–721 (2001).
- Basson, C. T. *et al.* Mutations in human *TBX5* cause limb and cardiac malformation in Holt-Oram syndrome. *Nature Genet.* **15**, 30–35 (1997).
- Li, Q. Y. *et al.* Holt-Oram syndrome is caused by mutations in *TBX5*, a member of the *Brachyury* (*T*) gene family. *Nature Genet.* **15**, 21–29 (1997).
- Davenport, T. G., Jerome-Majewska, L. A. & Papaioannou, V. E. Mammary gland, limb, and yolk sac defects in mice lacking *Tbx3*, the gene mutated in human ulnar mammary syndrome. *Development* **130**, 2263–2273 (2003).
- Agarwal, P. *et al.* *Tbx5* is essential for forelimb bud initiation following patterning of the limb field in the mouse embryo. *Development* **130**, 623–633 (2003).
- Hinchliffe, J. R. & Johnson, D. R. *The Development of the Vertebrate Limb: An Approach Through Experiment, Genetics, and Evolution* (Clarendon, Oxford, 1980).
- Martin, G. R. The roles of FGFs in the early development of vertebrate limbs. *Genes Dev.* **12**, 1571–1586 (1998).
- Johnson, R. L. & Tabin, C. J. Molecular models for vertebrate limb development. *Cell* **90**, 979–990 (1997).
- Niswander, L. Pattern formation: old models out on a limb. *Nature Rev. Genet.* **4**, 133–143 (2003).
- Saunders, J. W. Jr The proximo-distal sequence of the origin of the parts of the chick wing and the role of the ectoderm. *J. Exp. Zool.* **108**, 363–403 (1948).
- Barrow, J. *et al.* Ectodermal *Wnt3*/ β -catenin signaling is required for the establishment and maintenance of the apical ectodermal ridge. *Genes Dev.* **17**, 394–409 (2003).
- Fallon, J. *et al.* FGF-2: apical ectodermal ridge growth signal for chick limb development. *Science* **264**, 104–107 (1994).
- Niswander, L., Tickle, C., Vogel, A., Booth, I. & Martin, G. R. FGF-4 replaces the apical ectodermal ridge and directs outgrowth and patterning of the limb. *Cell* **75**, 579–587 (1993).
- Rowe, D. A., Cairns, J. M. & Fallon, J. F. Spatial and temporal patterns of cell death in limb bud mesoderm after apical ectodermal ridge removal. *Dev. Biol.* **93**, 83–91 (1982).
- Moon, A. M., Boulet, A. M. & Capecchi, M. R. Normal limb development in conditional mutants of *Fgf4*. *Development* **127**, 989–996 (2000).
- Sun, X. *et al.* Conditional inactivation of *Fgf4* reveals complexity of signalling during limb bud development. *Nature Genet.* **25**, 83–86 (2000).
- Saunders, J. W. Jr & Gasseling, M. T. In *Epithelial-Mesenchymal Interactions* (eds Fleischmajer, R. & Billingham, R. R.) 289–314 (Williams and Wilkins, Baltimore, 1968).
- Pearse, R. V. II & Tabin, C. J. The molecular ZPA. *J. Exp. Zool.* **282**, 677–690 (1998).
- Johnson, R. L., Riddle, R. D., Laufer, E. & Tabin, C. Sonic hedgehog: a key mediator of anterior-posterior patterning of the limb and dorso-ventral patterning of axial embryonic structures. *Biochem. Soc. Trans.* **22**, 569–574 (1994).
- Drossopoulou, G. *et al.* A model for anteroposterior patterning of the vertebrate limb based on sequential long- and short-range Shh signalling and Bmp signalling. *Development* **127**, 1337–1348 (2000).
- Lewis, P. M. *et al.* Cholesterol modification of Sonic hedgehog is required for long-range signaling activity and effective modulation of signaling by Ptc1. *Cell* **105**, 599–612 (2001).
- Litingtung, Y., Dahn, R. D., Li, Y., Fallon, J. F. & Chiang, C. *Shh* and *Gli3* are dispensable for limb skeleton formation but regulate digit number and identity. *Nature* **418**, 979–983 (2002).
- Aza-Blanc, P., Ramirez-Weber, F. A., Laget, M. P., Schwartz, C. & Kornberg, T. B. Proteolysis that is inhibited by Hedgehog targets Cubitus interruptus protein to the nucleus and converts it to a repressor. *Cell* **89**, 1043–1053 (1997).
- Chiang, C. *et al.* Manifestation of the limb prepattern: limb development in the absence of Sonic hedgehog function. *Dev. Biol.* **236**, 421–435 (2001).
- Kraus, P., Fraidenreich, D. & Loomis, C. A. Some distal limb structures develop in mice lacking Sonic hedgehog signaling. *Mech. Dev.* **100**, 45–58 (2001).
- te Welscher, P. *et al.* Progression of vertebrate limb development through SHH-mediated counteraction of GLI3. *Science* **298**, 827–830 (2002).
- Dahn, R. D. & Fallon, J. F. Interdigital regulation of digit identity and homeotic transformation by modulated BMP signaling. *Science* **289**, 438–441 (2000).
- Slack, J. M. W. *From Egg to Embryo: Determinative Events in Early Development* (Cambridge Univ. Press, Cambridge, 1983).
- Wolpert, L. Positional information and pattern formation. *Curr. Top. Dev. Biol.* **6**, 183–224 (1971).
- Summerbell, D., Lewis, J. H. & Wolpert, L. Positional information in chick limb morphogenesis. *Nature* **244**, 492–496 (1973).
- Wolpert, L., Lewis, J. & Summerbell, D. In *CIBA Foundation Symposium on Cell Patterning* 95–130 (Elsevier, London, 1974).
- Kieny, M. In *Vertebrate Limb and Somite Morphogenesis* (eds Ede, D. A., Hinchliffe, J. R. & Balls, M.) 87–103 (Cambridge Univ. Press, Cambridge, 1977).
- Kieny, M. & Pautou, M.-P. Proximo-distal pattern regulation in deficient avian limb buds. *Wilhelm Roux Arch.* **183**, 177–191 (1977).
- Hornbruch, A. & Wolpert, L. Cell division in the early growth and morphogenesis of the chick limb. *Nature* **226**, 764–766 (1970).
- Tickle, C. & Wolpert, L. The progress zone—alive or dead? *Nature Cell Biol.* **4**, E216–E217 (2002).
- Maini, P. K. & Solursh, M. Cellular mechanisms of pattern formation in the developing limb. *Int. Rev. Cytol.* **129**, 91–133 (1991).
- Kingsley, D. M. Genetic control of bone and joint formation. *Novartis Found. Symp.* **232**, 213–222 (2001).
- Wolpert, L. Limb patterning: reports of model's death exaggerated. *Curr. Biol.* **12**, R628–R630 (2002).

Acknowledgements We are grateful to J. Fallon, C. Tabin and colleagues in our laboratory for insightful criticism and suggestions for improving the text.

Cranial skeletal biology

J. A. Helms & R. A. Schneider

University of California at San Francisco, Room U-453, 533 Parnassus Avenue, San Francisco, California 94143-0514, USA
(e-mail: helms@itsa.ucsf.edu)

To artists, the face is a mirror of the soul. To biologists, the face reflects remarkable structural diversity — think of bulldogs and wolfhounds or galapagos finches. How do such variations in skeletal form arise? Do the same mechanisms control skeletogenesis elsewhere in the body? The answers lie in the molecular machinery that generates neural crest cells, controls their migration, and guides their differentiation to cartilage and bone.

"What is a face, really? Its own photo? Its make-up? Or is it a face as painted by such or such painter? ...Doesn't everyone look at himself in his own particular way? Deformations simply do not exist." Pablo Picasso¹

Despite its remarkable topological complexity, one might legitimately wonder whether studying skeletal development in the head is a little superfluous. Is there any reason to suspect that factors controlling chondrogenesis and osteogenesis in the head are distinguishable from those operating in the limbs or the spine? If skeletal development in the head simply parallels skeletogenesis elsewhere in the body, then studying this process in such a complicated structure would seem to be a gratuitous endeavour. As it turns out, the molecular mechanisms inducing chondrogenesis and osteogenesis in cranial neural crest cells, which produce the facial and jaw skeleton, are distinct from those operating in mesodermal cells, which produce the remainder of the skeleton. Our ability to prevent or at least mitigate cranial skeletal anomalies, and to enhance cranial skeletal repair, depends upon understanding these cranial-specific pathways.

Organization of the cranial skeleton

For those interested in skeletal biology, the appendicular and axial skeletons hold a distinct appeal for analysis, as their elements exhibit an uncomplicated anatomy. Regardless of whether an animal uses its forelimbs for flying or playing flamenco guitar, the basic morphology is preserved. Likewise, axial organization is highly conserved: even mammals as dissimilar as a giraffe and a porpoise possess the same number of cervical vertebrae. The cranial skeleton, on the other hand, is composed of an often-bewildering assortment of neural crest- and mesoderm-derived cartilages and bones that have been highly modified during evolution. This makes comparisons among divergent species a challenge.

Such anatomical complexity and embryonic amalgamations may initially diminish one's enthusiasm for understanding how the cranial skeleton is generated and undergoes repair. Nonetheless, studies in cranial skeletogenesis have provided unparalleled insights into the cellular and molecular mechanisms that generate and shape cartilage and bone. Our growing realization that skeletogenesis in the head is a unique and separable process from that occurring elsewhere in the body makes this a difficult, but worthwhile, road to take.

Genesis and exodus of the cranial neural crest

Most of the head skeleton is derived from the cranial neural crest, which arises from the dorsal margins of the

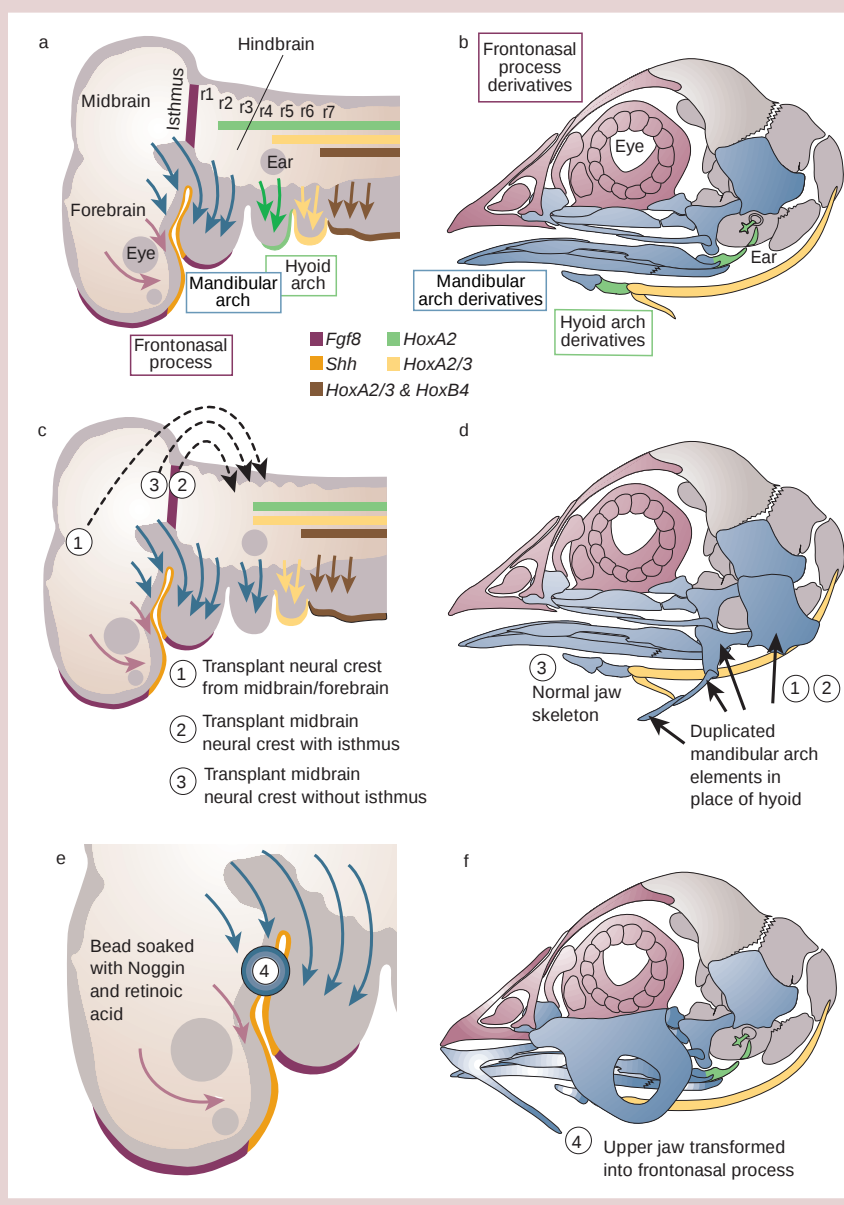
neural folds. Bronner-Fraser and colleagues showed that a Wnt protein is necessary and sufficient for neural crest induction². In the absence of Wnt signalling, neural crest cells were not generated. Exogenous Wnt protein was sufficient to regenerate the missing crest, and could induce neural crest from naive neural ectoderm. New data indicate that Snail family members are also involved in the generation and early migration of neural crest³, but how they interact with Wnts and bone morphogenetic proteins (BMPs) remains a mystery.

The last century has brought an appreciation for how cranial neural crest cells get from 'here', the dorsal part of the neural tube, to 'there', the facial primordia. Whereas coal dust was once sprinkled over avian embryos to label migrating neural crest cells, innovative visualization techniques now afford us with a bird's eye view of migratory routes and cellular events that occur during neural crest emigration from the neural tube⁴. Why was there such interest in the movement of neural crest cells? The consequences of inaccurate or interrupted migration cannot be over-emphasized, as a plethora of craniofacial malformations have as their primary aetiology a perturbation in this cellular exodus. For example, pharyngeal arch syndromes and neurocristopathies are associated with aberrant neural crest migration^{5,6}.

Intravital imaging movies now allow investigators to follow neural cell migration in real time and, in so doing, have revealed unexpected cell behaviours⁷. Neural crest cells do not meander on their journey to the craniofacial primordia. Instead, they display highly stereotyped migratory patterns, passing between neural and facial epithelia, and around paraxial mesoderm until they reach their locations in the pharyngeal arches and frontonasal process. The course is not hard-wired within a given neural crest population. Cells transplanted to ectopic locations take their cues from the local environment and arrive in the position that corresponds to their new level of origin^{4,8}. These studies offer fresh insights into aetiologies of neurocristopathies affecting the facial skeleton.

Signals encountered by neural crest cells during their migration to the facial primordia can alter their fate. Some guidance cues have been identified and a common feature among them is that they already have well-documented roles in axon path finding^{9–11}. Experiments conducted by Golding, Gassmann and colleagues demonstrate this point. The receptor tyrosine kinase ErbB4 is involved in controlling neural crest cell migration^{12,13}, but ErbB4 is not expressed on

Figure 1 Experimental approaches to skeletal patterning. **a**, Cellular and molecular organization of a developing avian head shown as a schematized composite of several embryonic stages that illustrates (in lateral view) anatomical relations among the developing brain, migrating neural crest cells (arrows), sensory structures and craniofacial primordia (frontonasal process, mandibular arch and hyoid arch). *Hox* genes show periodic expression in hindbrain rhombomeres (*r*). *Fgf8* and *Shh* are expressed at the midbrain/hindbrain boundary (isthmus) and in epithelia surrounding the craniofacial primordia. Drawing modified from ref. 37. **b**, The facial skeleton arises from neural crest cells that migrate into the frontonasal process, mandibular arch and hyoid arch. Drawing modified from ref. 78. **c**, Several experimental approaches have been taken to test whether neural crest cells are pre-patterned. These include, replacement of hyoid arch neural crest with cells from the midbrain/forebrain boundary (1) and the midbrain/hindbrain boundary (including the isthmus) (2)²⁹, and from the midbrain/hindbrain boundary (without the isthmus) (3)³¹. **d**, Experimental results (1 and 2) include duplication of some elements of the mandibular arch as well as formation of a normal jaw skeleton (3). **e**, Another experimental approach (4) uses beads soaked in specific molecules (such as Noggin and retinoic acid), which alter the neural crest identity⁷⁹. **f**, Results include the transformation of maxillary elements into frontonasal process structures.



neural crest cells. Rather, the protein is abundantly expressed in neural ectoderm¹⁴ and thus constitutes a kind of 'roadside cue' to migrating cells. These investigators surmised that the misplaced axons they observed in *ErbB4*^{-/-} mice might be harbingers of a more global disorganization in cell migration, which occurs in the absence of *ErbB4*. By culturing embryos *in vitro*, the group showed that *ErbB4*^{-/-} cells transplanted into a wild-type host embryo migrated normally, whereas wild-type cells placed into a mutant host deviated from their prescribed route. These data are compelling because they provide direct evidence that epithelia can supervise and instruct migration of neural crest cells, and that there is a shared mechanism for guiding axons and the migrating neural crest.

Skeletogenic capacity of the cranial neural crest

Cranial neural crest cells possess an ability to form cartilage and bone, whereas trunk neural crest cells have minimal skeletogenic capacity¹⁵. Precisely why trunk neural crest cells do not contribute to the axial or appendicular skeleton is somewhat of a conundrum. Is it because trunk neural crest cells are not exposed to appropriate 'skeletogenic' cues in their environment, or because they are

somehow restricted in their competency to form skeletal tissue? If the cranial environment is important for skeletogenic differentiation, then transplanting trunk neural crest to the head should allow trunk cells to form cartilage or bone. But regardless of the axial level to which they were transplanted, trunk neural crest cells failed to make skeletal tissues¹⁶. Even when trunk neural crest cells were explanted and treated with BMPs^{17,18}, which promote skeletogenesis in other tissues, the cells failed to differentiate like their cranial counterparts^{19,20}. Other investigators showed that trunk neural crest cells, if grown for extended periods of time *in vitro*, could differentiate into chondrocytes¹⁵. Likewise, when small grafts of trunk neural crest were transplanted into the head region, a few cells migrated to the appropriate axial level and contributed to cranial cartilages²¹.

How does one reconcile this experimental paradox? One possibility is that a 'community effect' influences the skeletogenic capacity of trunk neural crest. When trunk cells are sufficiently dispersed, inhibitory effects from other trunk cells may be mitigated and a few cells may then respond to skeletogenic cues in their environment. Although a small group of cells might hear these signals, a large

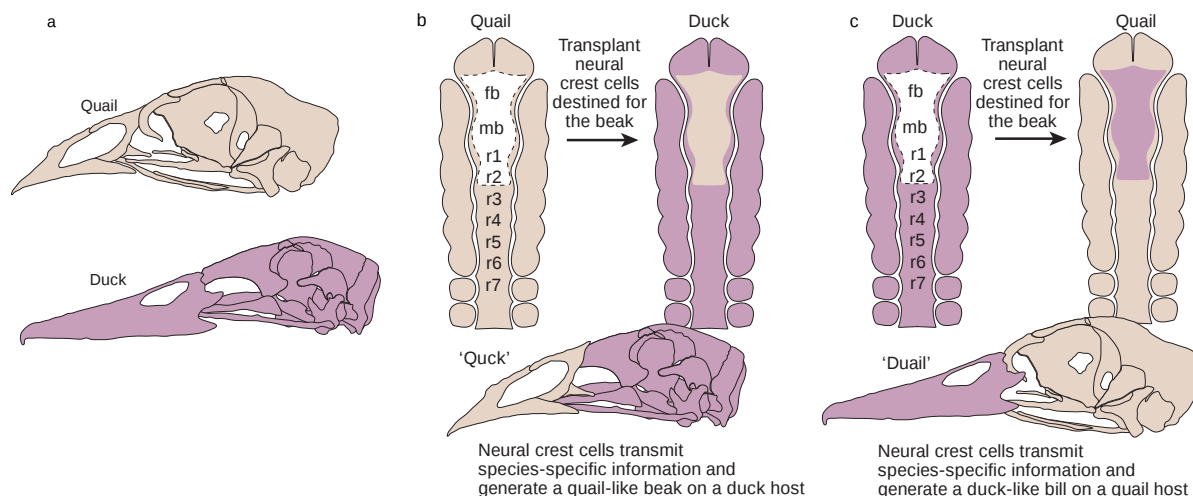


Figure 2 Contributions of neural crest to beak morphology. **a**, Quail and duck embryos are anatomically distinct and thus are a powerful system for studying craniofacial patterning. **b**, When quail neural crest cells from forebrain (fb), midbrain (mb) and first

two rhombomeres (r) are transplanted into a duck host, a quail-like beak forms instead of a duck's bill³⁵. **c**, Likewise, duck neural crest cells make a duck-like bill when transplanted into a quail host. Drawings modified from refs 80, 81.

population of trunk neural crest may be impervious. What actually constitutes this inhibitory signal is still unknown. In any case, gauging the skeletogenic capacity of different neural crest populations is of vital interest to those who seek to stimulate regeneration of cranial skeletal tissues.

A dual origin of the cranial skeleton

Twenty years ago, the 'new head' hypothesis proposed that much of the cranial skeleton was formed by expansion of a rostral neural crest cell population, and that the dividing line between mesoderm and neural crest had significance for the development and evolution of the skeleton²². A genetic approach provides convincing new data about the relative contributions of neural crest and mesoderm to the cranial skeleton. *Wnt1* is expressed in the dorsal neural tube, around the time of neural crest induction. By placing the *Wnt1* promoter upstream of the *Cre* gene and crossing mice carrying this transgene with a reporter line, indelibly labelled neural crest cells were generated whose ultimate fate could be followed throughout the lifetime of the animal²³.

These data supported earlier avian fate maps showing the dual neural crest and mesodermal origin of the cranial vault²⁴, and underscored previous observations²⁵ that the neural crest cells contribute to another cranial skeletal tissue, the teeth. Dentin, dental pulp, alveolar bone and the periodontal ligament, which anchors the teeth to the bone of the jaws, are all derived from the neural crest²⁶. These studies are a necessary first step towards using undifferentiated adult neural crest cells from the pulp and marrow cavity for the repair of dental and skeletal defects. Patients cursed with cavities and gum disease will undoubtedly take great comfort from the fact that cellular and molecular therapies, rather than drills and scalpels, may one day be used by dentists to treat their afflictions^{27,28}.

Establishment of cranial skeletal architecture

After more than a century of exploring mechanisms that generate the cranial skeleton, two prevailing theories have emerged. The first presupposes that cranial neural crest cells carry out autonomous programs for patterning, while the second theory posits that neural crest cells are naive and acquire patterning information through interactions with the local environment.

Almost 20 years ago, Noden showed that moving neural crest cells from the midbrain to a more posterior location in the hindbrain resulted in an embryo with a duplicated jaw skeleton²⁹. These data suggested that neural crest cells contained, at the time of their transplantation, intrinsic information for elaborating the complex morphology of the jaw skeleton (Fig. 1c, d). This interpretation was challenged by Le Douarin and Couly³⁰, and more recently by Krumlauf and colleagues, who suspected that the duplicated jaw skeleton arose as a consequence of Noden inadvertently transplanting an adjacent region of the neural tube called the isthmus³¹. The isthmus expresses *fibroblast growth factor 8* (*Fgf8*), a gene encoding a secreted protein that can downregulate the homeobox gene *HoxA2*, whose own expression is required for the formation of second arch structures^{32–34}. If Noden's graft included isthmus tissue then *HoxA2* should be downregulated and the result would be duplicated lower jaw structures. When midbrain/hindbrain crest were transplanted without the isthmus the result was a normal jaw skeleton, and when the transplants included the isthmus the result was a duplicated jaw skeleton.

These data suggested that neural crest cells do not possess inherent patterning information. Or do they? When we exchanged neural crest cells of the presumptive beak region between quail and duck embryos, the facial features of the chimaeras more closely resembled the donor species³⁵. When quail neural crest cells destined to form the beak were transplanted to a duck host, the result was a duck embryo with a quail-like beak ('quack'). Conversely, when duck cells were transplanted into quail hosts, quail embryos with duck-like bills ('duails') were produced (Fig. 2). Molecular and cellular analyses revealed that donor neural crest cells followed their own morphogenetic program and re-patterned host facial ectoderm like that of the donor species. These data demonstrate that neural crest cells can influence surrounding tissues.

To what extent do surrounding tissues regulate neural crest cell fate? Hu, Marcucio and Helms showed that a zone of frontonasal ectoderm stimulated the proliferation and differentiation of underlying neural crest cells (Fig. 3). When transplanted to an ectopic location, the frontonasal ectodermal zone activated a cascade of molecular events that ultimately re-programmed the neural crest-derived mesenchyme and produced a duplication in

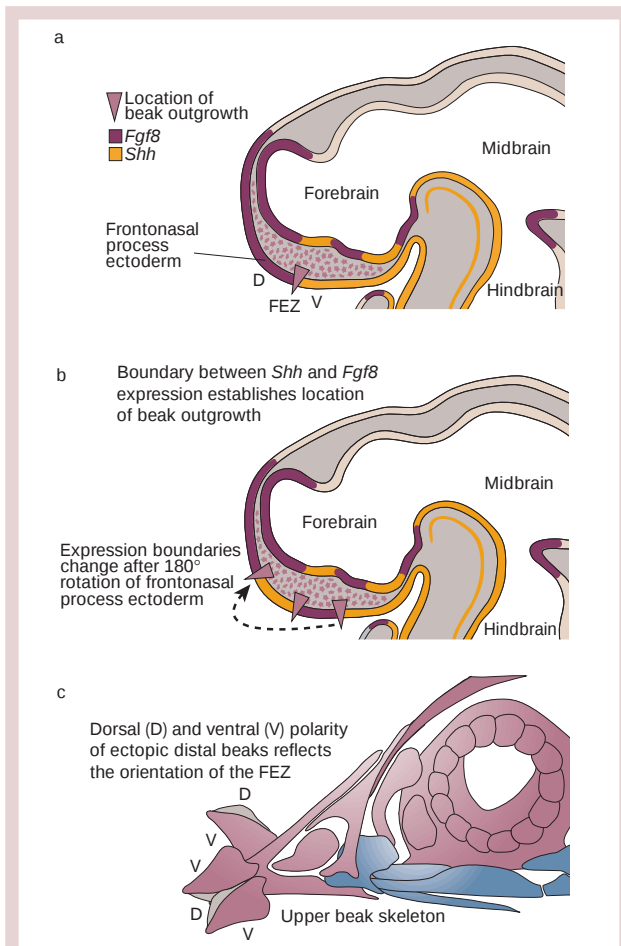


Figure 3 Contributions of epithelia to craniofacial patterning. **a**, A molecular boundary in frontonasal ectoderm, defined by *Fgf8* and *Shh* expression, presages the initial site of outgrowth (arrowhead) and defines dorsoventral polarity of the upper beak³⁶. **b**, Changing the dorsoventral orientation of the frontonasal ectodermal zone (FEZ) alters dorsoventral orientation of the upper beak. A 180° rotation results in three sites where dorsal and ventral ectodermal domains are juxtaposed (arrowheads). **c**, As a result, distal beak structures are duplicated. The first (dorsal-most) beak has a dorsal-ventral (DV) pattern, whereas the second beak has a ventral-dorsal-ventral (VDV) pattern, indicating the ability of epithelium to re-pattern skeletal elements derived from the neural crest.

upper beak structures³⁶. Other epithelia in the head, such as the pharyngeal endoderm, also influence the size, shape and position of some components of the facial skeleton³⁷ (Box 1).

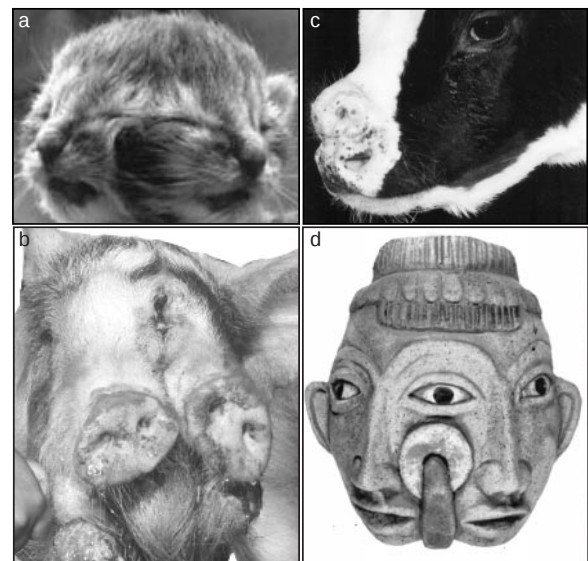
Craniofacial defects provide a window into morphogenesis

Higher vertebrates have evolved specialized neural mechanisms for the recognition of faces³⁸, providing us with an amazing ability to discriminate among hundreds of people. This same exquisite tuning permits us to detect even subtle discrepancies in facial form. Craniofacial malformations compromise not only function (for example, speech and mastication) but also exact a demoralizing toll on the mental well being of the affected individual. Recent advances in human genetics and experimental embryology indicate that we are converging on an understanding of how particular gene perturbations produce cranial skeletal malformations. The zebrafish has become a crucial model system for exploring these questions (reviewed in ref. 39).

Lineage studies have defined the origins of the zebrafish cranial skeleton at the single-cell level⁴⁰, and molecular analyses indicate that the same nested expression patterns of homeobox genes that regulate

Box 1 Facial duplications

We take for granted that faces have two eyes, a nose and a mouth. What happens when nature produces an animal with three or four eyes, two noses and two mouths? These duplications, and those at the opposite end of the spectrum such as cyclopia, stretch our imagination and provoke us to understand the molecular and cellular events that had to have occurred to produce such phenomenal phenotypes. In the figure below, 'Image' the cat (panel **a**) and 'Ditto' the pig (panel **b**) are examples not of twinning, but of actual craniofacial duplications. Other duplications can be limited to individual facial structures, such as that seen on the nose of the calf (panel **c**; from D. Noden). Facial duplications are documented throughout mythology. The ancient mask in panel **d** represents a god who possesses the ability to speak truthfully and deceitfully, as well as see the future, present and past (from Z. Werb).



skeletal patterning in birds⁴¹ and mammals^{42,43} also participate in zebrafish head morphogenesis⁴⁴. As might be suspected from such conserved patterns of homeobox gene expression, mutations in mice and knockdowns in fish produce remarkably similar phenotypes. Disruptions in a Hox binding partner, *lazarus/Pbx4*^{45,46}, and a Hox gene regulator, *valentino/Kreisler*⁴⁷, produce malformations in the pharyngeal cartilages. A knockdown in zebrafish Hox2 function produces defects in second pharyngeal arch cartilage, where ventral skeletal elements were replaced by duplicated first arch structures⁴⁸, analogous to the murine *HoxA2*^{-/-} phenotype^{32,33}.

Precisely how these disruptions influence cellular interactions and produce skeletal defects is not clear, but analyses of *sucker/Endothelin-1* (*suc/Et-1*) gene function have provided some exciting clues⁴⁹. The zebrafish gene *suc*, and the amniote gene *Et-1*, encode a secreted peptide^{50,51}. In the pharyngeal arches *Et-1* is expressed in a core of mesoderm surrounded by a sheet of *Et-1*-negative neural crest mesenchyme, which in turn is wrapped in *Et-1*-positive epithelia⁴⁹. Loss of *suc/Et-1* disrupts gene expression in the ventral but not the dorsal portion of the pharyngeal arches, which suggests that there is a morphogenetic subdivision of the pharyngeal arch skeleton⁵². Mosaic analyses also indicate that a *suc/Et-1* signal from mesoderm and/or epithelia may be required for a subgroup of neural crest cells to adopt a skeletogenic fate⁵². Thus, interactions between epithelia, mesoderm and neural crest-derived cells are required for the initial steps of skeletogenesis.

Sometimes the clearest view of normal development is attained when embryonic processes go awry. Teratogens offer a window into fetal development, but until recently we had little insight into mechanisms by which these agents exert their untoward effects. Fetal alcohol syndrome is an entirely avoidable suite of physical, mental and neurobehavioural birth defects caused by alcohol consumption during pregnancy. Although it has been appreciated for many years that alcohol-related malformations are associated with excessive cell death, Ahlgren and Bronner-Fraser demonstrated that this cell death occurs via a Sonic hedgehog (Shh)-dependent mechanism⁵³. How are the teratogenic effects of alcohol and the signalling functions of Shh connected? Hu and Helms showed that Shh inhibition in the face caused the downregulation of Patched and Gli1, and induced hypotelorism and facial clefting⁵⁴. Another molecule, retinoic acid (a derivative of vitamin A) is essential for life but, in surplus or shortage, acts as a teratogen. Alcohol and retinol are metabolized by closely related enzymes⁵⁵, which suggests that deficiencies in either molecule could disrupt craniofacial morphogenesis along similar routes. Schneider, Hu and Helms showed that blocking retinoid signalling in the rostral head leads to a loss of Fgf8 and Shh and their downstream effectors⁵⁶, and causes morphological defects reminiscent of those induced by alcohol and Shh inhibition.

Another teratogen, cyclopamine, produces holoprosencephalic defects in mammals⁵⁷ and avians⁵⁸ (reviewed in ref. 59). Cyclopamine acts specifically by inhibiting Shh signalling⁶⁰. The teratogenic consequences of cyclopamine exposure depend upon the gestational age of the fetus. If embryos are exposed during neurulation, cyclopic defects are observed⁶⁰. If embryos are exposed at later developmental stages, the results range from premaxillary agenesis and cleft lip/palate to other midline anomalies⁶¹. A similar spectrum of craniofacial anomalies are seen in *Shh*^{-/-} embryos, which are cyclopic⁶² and *Cdon*^{-/-} mice, which exhibit subtle midline abnormalities such as the loss of a central incisor⁶³.

In some cases the deletion of a gene does not truncate, but instead transforms facial growth. Such is the case for mice lacking Distal-less homeobox gene 5 (*Dlx5*) and *Dlx6*. The mandibular primordia were transformed into ectopic maxillary primordia, complete with epidermal elaborations and skeletal tissue duplications unique to the upper jaw^{64,65}.

Clearly then, a dialogue exists among multiple tissues that mediates cranial skeletogenesis. Changes in this 'molecular conversation' can alter craniofacial morphology in astonishing ways, and the combination of genetic techniques with epigenetic approaches will prove to be a valuable tool in unravelling the regulation of craniofacial morphogenesis.

Bone formation gone wrong

During the past decade there has been a dramatic increase in our understanding of the molecular underpinnings of craniosynostoses⁶⁶. We now appreciate that activating mutations in FGF receptors are responsible for a number of craniosynostotic conditions⁶⁷. One puzzling aspect is that FGFs and their receptors are widely expressed during fetal development, yet mutations produce localized skeletal defects. Insight into this issue has come from a recent study addressing molecular differences between fusing and non-fusing (patent) sutures. Signals from the dura mater regulate osteogenesis in overlying skeletogenic mesenchyme⁶⁸, and likely include BMP-family members. In mice, FGFs and BMPs are widely expressed in suture mesenchyme, but only some of the sutures fuse. What regulates the site-specific activity of these growth factors? Longaker, Harland and colleagues showed that the BMP antagonist Noggin was expressed in all sutures, but its expression was downregulated by Fgf2 only in sutures that fuse⁶⁹. Moreover, ectopic Fgf2 expression in a non-fusing suture led to Noggin repression, which caused open sutures to undergo fusion, whereas Noggin mis-expression caused fusing sutures to remain open. The group further demonstrated that activating mutations in Fgf2 diminished

Noggin expression, raising the possibility that premature suture fusion might one day be treated by Noggin.

The sites where skull roofing bones coalesce (that is, the sutures) are presaged by the expression of muscle segment homeobox (*Msx*) transcription factors⁷⁰. As might be imagined, disruptions in *Msx* genes create ossification defects in the skull bones of mice⁷¹ and humans⁷². The severity of these skull defects is inversely proportional to *Msx* gene dosage⁷¹. In an effort to understand how mutations in *Msx2* produce such skeletal phenotypes, Maxson and co-workers created *Msx2* deletions in *Wnt1*-Cre/R26R mice, and showed that the number of *LacZ*-positive cells in the skeletogenic condensations was equivalent between wild-type and mutant embryos. Thus, the skull bone defects were not due to disruptions in the migration or general allocation of neural crest cells to the frontal bone rudiment (R. Maxson, personal communication). Instead, *Sox9* expression and alkaline phosphatase activity were reduced in the mutant mice, indicating that the skeletal defects were caused because of a failure in the specification or proliferation of skeletogenic mesenchyme. This phenotype bears a striking resemblance to the zebrafish mutant *suc/Et-1* described earlier, which fails to develop ventral pharyngeal cartilages because of a defect in specification of skeletogenic mesenchyme. *MsxE* expression is lost in *suc/Et-1* mutants⁵², lending further support for an epistatic relationship between these molecules.

The final frontier for cranial tissue regeneration

The burgeoning field of stem cell biology is an unambiguous indicator of a scientific interest in the regenerative potential of the human body. There are significant gaps in our understanding of stem cell biology that will need to be addressed before therapeutic strategies can be routinely used. For example, the origins of stem cells that form regenerating bone have been surmised but not proven. Influences from the microenvironment in which stem cells reside are also poorly defined. The molecular signals and mechanical stimuli that induce cells to differentiate still remain largely unknown⁷³⁻⁷⁶.

As techniques to diagnose cranial skeletal defects become increasingly reliable, we are confronted with the possibility of treating milder forms of skeletal malformations in the fetus⁷⁷, which would obviate the need for numerous postnatal surgeries. Is our basic science knowledge sufficient to guide surgeons in this endeavour? Most of our information on bone repair comes from analyses of long bone fractures, but accumulating evidence suggests that cranial skeletal repair is a distinctive process. Identifying a source of cranial skeletogenic stem cells, elucidating the molecular signals that drive their differentiation, and developing delivery methods for these cells are worthy objectives. Such research will have direct and profound implications for the treatment of cranial skeletal defects resulting from malformation, disease or injury. □

doi:10.1038/nature01656

- Picasso, P. *Picasso on Art: A Selection of Views* (ed. Dore, A.) (Thames and Hudson, London, 1972).
- Garcia-Castro, M. I., Marcelle, C. & Bronner-Fraser, M. Ectodermal Wnt function as a neural crest inducer. *Science* **13**, 13 (2002).
- Aybar, M. J., Nieto, M. A. & Mayor, R. *Snail* precedes *slug* in the genetic cascade required for the specification and migration of the *Xenopus* neural crest. *Development* **130**, 483–494 (2003).
- Birgbauer, E., Sechrist, J., Bronner-Fraser, M. & Fraser, S. Rhombomeric origin and rostrocaudal reassignment of neural crest cells revealed by intravital microscopy. *Development* **121**, 935–945 (1995).
- Vitelli, F., Morishima, M., Taddei, L., Lindsay, E. A. & Baldini, A. *Tbx1* mutation causes multiple cardiovascular defects and disrupts neural crest and cranial nerve migratory pathways. *Hum. Mol. Genet.* **11**, 915–922 (2002).
- Garg, V. et al. *Tbx1*, a DiGeorge syndrome candidate gene, is regulated by Sonic hedgehog during pharyngeal arch development. *Dev. Biol.* **235**, 62–73 (2001).
- Kulesa, P. M. & Fraser, S. E. In ovo time-lapse analysis of chick hindbrain neural crest cell migration shows cell interactions during migration to the branchial arches. *Development* **127**, 1161–1172 (2000).
- Kulesa, P., Bronner-Fraser, M. & Fraser, S. In ovo time-lapse analysis after dorsal neural tube ablation shows rerouting of chick hindbrain neural crest. *Development* **127**, 2843–2852 (2000).
- Holder, N. & Klein, R. Eph receptors and ephrins: effectors of morphogenesis. *Development* **126**, 2033–2044 (1999).
- Smith, A., Robinson, V., Patel, K. & Wilkinson, D. G. The EphA4 and EphB1 receptor tyrosine kinases and ephrin-B2 ligand regulate targeted migration of branchial neural crest cells. *Curr. Biol.* **7**, 561–570 (1997).
- Eickholt, B. J., Mackenzie, S. L., Graham, A., Walsh, F. S. & Doherty, P. Evidence for collapsin-1

- functioning in the control of neural crest migration in both trunk and hindbrain regions. *Development* **126**, 2181–2189 (1999).
12. Golding, J. P., Trainor, P., Krumlauf, R. & Gassmann, M. Defects in pathfinding by cranial neural crest cells in mice lacking the neuregulin receptor ErbB4. *Nature Cell Biol.* **2**, 103–109 (2000).
13. Gassmann, M. *et al.* Aberrant neural and cardiac development in mice lacking the ErbB4 neuregulin receptor. *Nature* **378**, 390–394 (1995).
14. Golding, J. P., Tidcombe, H., Tsoni, S. & Gassmann, M. Chondroitin sulphate-binding molecules may pattern central projections of sensory axons within the cranial mesenchyme of the developing mouse. *Dev. Biol.* **216**, 85–97 (1999).
15. McConnell, I. M. & Graham, A. Trunk neural crest has skeletogenic potential. *Curr. Biol.* **12**, 767–771 (2002).
16. Nakamura, H. & Ayer-le Lievre, C. S. Mesodermal capabilities of the trunk neural crest of birds. *J. Embryol. Exp. Morphol.* **70**, 1–18 (1982).
17. Shah, N. M., Groves, A. K. & Anderson, D. J. Alternative neural crest cell fates are instructively promoted by TGF β superfamily members. *Cell* **85**, 331–343 (1996).
18. Smith, A. & Graham, A. Restricting Bmp-4 mediated apoptosis in hindbrain neural crest. *Dev. Dyn.* **220**, 276–283 (2001).
19. Anderson, D. J. *et al.* Cell lineage determination and the control of neuronal identity in the neural crest. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 493–504 (1997).
20. Baroffio, A., Dupin, E. & Le Douarin, N. M. Clone-forming ability and differentiation potential of migratory neural crest cells. *Proc. Natl Acad. Sci. USA* **85**, 5325–5329 (1988).
21. Groves, A. K. & Bronner-Fraser, M. Competence, specification and commitment in otic placode induction. *Development* **127**, 3489–3499 (2000).
22. Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates: a new head. *Science* **220**, 268–274 (1983).
23. Jiang, X., Iseki, S., Maxson, R. E., Sucov, H. M. & Morriss-Kay, G. M. Tissue origins and interactions in the mammalian skull vault. *Dev. Biol.* **241**, 106–116 (2002).
24. Noden, D. M. The control of avian cephalic neural crest cytodifferentiation. I. Skeletal and connective tissues. *Dev. Biol.* **67**, 296–312 (1978).
25. Lumsden, A. G. Spatial organization of the epithelium and the role of neural crest cells in the initiation of the mammalian tooth germ. *Development* **103**(Suppl.), 155–169 (1988).
26. Chai, Y. *et al.* Fate of the mammalian cranial neural crest during tooth and mandibular morphogenesis. *Development* **127**, 1671–1679 (2000).
27. Chai, Y. & Slavkin, H. C. Prospects for tooth regeneration in the 21st century: a perspective. *Microsc. Res. Tech.* **60**, 469–479 (2003).
28. Wilson, C. Cutting edge. *New Sci.* **175**, 32 (2002).
29. Noden, D. M. The role of the neural crest in patterning of avian cranial skeletal, connective, and muscle tissues. *Dev. Biol.* **96**, 144–165 (1983).
30. Couly, G., Grapin-Botton, A., Coltey, P., Ruhin, B. & Le Douarin, N. M. Determination of the identity of the derivatives of the cephalic neural crest: incompatibility between *Hox* gene expression and lower jaw development. *Development* **125**, 3445–3459 (1998).
31. Trainor, P. A., Ariza-McNaughton, L. & Krumlauf, R. Role of the isthmus and FGFs in resolving the paradox of neural crest plasticity and prepattern. *Science* **295**, 1288–1291 (2002).
32. Gendron-Maguire, M., Mallo, M., Zhang, M. & Gridley, T. Hoxa-2 mutant mice exhibit homeotic transformation of skeletal elements derived from cranial neural crest. *Cell* **75**, 1317–1331 (1993).
33. Rijli, F. M. *et al.* A homeotic transformation is generated in the rostral branchial region of the head by disruption of *Hoxa-2*, which acts as a selector gene. *Cell* **75**, 1333–1349 (1993).
34. Creuzet, S., Couly, G., Vincent, C. & Le Douarin, N. M. Negative effect of *Hox* gene expression on the development of the neural crest-derived facial skeleton. *Development* **129**, 4301–4313 (2002).
35. Schneider, R. A. & Helms, J. A. The cellular and molecular origins of beak morphology. *Science* **299**, 565–568 (2003).
36. Hu, D., Marcucio, R. & Helms, J. A. A zone of frontonasal ectoderm regulates patterning and growth in the face. *Development* **130**, 1749–1758 (2003).
37. Couly, G., Creuzet, S., Bannaceur, S., Vincent, C. & Le Douarin, N. M. Interactions between *Hox*-negative cephalic neural crest cells and the foregut endoderm in patterning the facial skeleton in the vertebrate head. *Development* **129**, 1061–1073 (2002).
38. Kendrick, K. M., da Costa, P., Leigh, A. E., Hinton, M. R. & Peirce, J. W. Sheep don't forget a face. *Nature* **414**, 165–166 (2001).
39. Kimmel, C. B., Miller, C. T. & Moens, C. B. Specification and morphogenesis of the zebrafish larval head skeleton. *Dev. Biol.* **233**, 239–257 (2001).
40. Schilling, T. F. & Kimmel, C. B. Segment and cell type lineage restrictions during pharyngeal arch development in the zebrafish embryo. *Development* **120**, 483–494 (1994).
41. Prince, V. & Lumsden, A. *Hoxa-2* expression in normal and transposed rhombomeres: independent regulation in the neural tube and neural crest. *Development* **120**, 911–923 (1994).
42. Trainor, P. A. & Krumlauf, R. Patterning the cranial neural crest: hindbrain segmentation and *Hox* gene plasticity. *Nature Rev. Neurosci.* **1**, 116–124 (2000).
43. Capecchi, M. R. *Hox* genes and mammalian development. *Cold Spring Harb. Symp. Quant. Biol.* **62**, 273–281 (1997).
44. Amores, A. *et al.* Zebrafish *hox* clusters and vertebrate genome evolution. *Science* **282**, 1711–1714 (1998).
45. Popperl, H. *et al.* *lazarus* is a novel *pbx* gene that globally mediates *hox* gene function in zebrafish. *Mol. Cell* **6**, 255–267 (2000).
46. Cooper, K. L., Leisenring, W. M. & Moens, C. B. Autonomous and nonautonomous functions for *Hox/Pbx* in branchiomotor neuron development. *Dev. Biol.* **253**, 200–213 (2003).
47. Prince, V. E., Moens, C. B., Kimmel, C. B. & Ho, R. K. Zebrafish *hox* genes: expression in the hindbrain region of wild-type and mutants of the segmentation gene, *valentino*. *Development* **125**, 393–406 (1998).
48. Hunter, M. P. & Prince, V. E. Zebrafish *Hox* paralogue group 2 genes function redundantly as selector genes to pattern the second pharyngeal arch. *Dev. Biol.* **247**, 367–389 (2002).
49. Schilling, T. F. *et al.* Jaw and branchial arch mutants in zebrafish I: branchial arches. *Development* **123**, 329–344 (1996).
50. Yanagisawa, M. *et al.* A novel potent vasoconstrictor peptide produced by vascular endothelial cells. *Nature* **332**, 411–415 (1988).
51. Remuzzi, G., Perico, N. & Benigni, A. New therapeutics that antagonize endothelin: promises and frustrations. *Nature Rev. Drug Discov.* **1**, 986–1001 (2002).
52. Miller, C. T., Schilling, T. F., Lee, K., Parker, J. & Kimmel, C. B. *sucker* encodes a zebrafish Endothelin-1 required for ventral pharyngeal arch development. *Development* **127**, 3815–3828 (2000).
53. Ahlgren, S. C., Thakur, V. & Bronner-Fraser, M. Sonic hedgehog rescues cranial neural crest from cell death induced by ethanol exposure. *Proc. Natl Acad. Sci. USA* **99**, 10476–10481 (2002).
54. Hu, D. & Helms, J. A. The role of Sonic hedgehog in normal and abnormal craniofacial morphogenesis. *Development* **126**, 4873–4884 (1999).
55. Napoli, J. L. Interactions of retinoid binding proteins and enzymes in retinoid metabolism. *Biochim. Biophys. Acta* **1440**, 139–162 (1999).
56. Schneider, R. A., Hu, D., Rubenstein, J. L., Maden, M. & Helms, J. A. Local retinoid signaling coordinates forebrain and facial morphogenesis by maintaining FGF8 and SHH. *Development* **128**, 2755–2767 (2001).
57. Keeler, R. F. & Binns, W. Teratogenic compounds of *Veratrum californicum* (Durand). V. Comparison of cyclopian effects of steroidal alkaloids from the plant and structurally related compounds from other sources. *Teratology* **1**, 5–10 (1968).
58. Incardona, J. P., Gaffield, W., Kapur, R. P. & Roelink, H. The teratogenic *Veratrum* alkaloid cyclopamine inhibits Sonic hedgehog signal transduction. *Development* **125**, 3553–3562 (1998).
59. Cordero, D. R., Schneider, R. A. & Helms, J. A. in *Craniofacial Surgery: Science & Surgical Technique* (eds Lin, K. Y., Ogle, R. C. & Jane, J. A.) 75–83 (W. B. Saunders, Philadelphia, 2002).
60. Cooper, M. K., Porter, J. A., Young, K. E. & Beachy, P. A. Teratogen-mediated inhibition of target tissue response to Shh signaling. *Science* **280**, 1603–1607 (1998).
61. Cordero, D. R., Marcucio, R., Gaffield, W., Tapadia, M. & Helms, J. A. Temporal disruption in Sonic Hedgehog signalling mimics the phenotypic range of holoprosencephaly. *Nature Med.* (submitted).
62. Chiang, C. *et al.* Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature* **383**, 407–413 (1996).
63. Cole, F. & Krauss, R. S. Microform holoprosencephaly in mice that lack the Ig superfamily member Cdon. *Curr. Biol.* **13**, 411–415 (2003).
64. Beverdam, A. *et al.* Jaw transformation with gain of symmetry after *Dlx5/Dlx6* inactivation: mirror of the past? *Genesis* **34**, 221–227 (2002).
65. Depew, M. J., Lufkin, T. & Rubenstein, J. L. Specification of jaw subdivisions by *Dlx* genes. *Science* **298**, 381–385 (2002).
66. Cohen, M. M. Jr Craniofacial disorders caused by mutations in homeobox genes *MSX1* and *MSX2*. *J. Craniofac. Genet. Dev. Biol.* **20**, 19–25 (2000).
67. Gorlin, R. J. Fibroblast growth factors, their receptors and receptor disorders. *J. Craniomaxillofac. Surg.* **25**, 69–79 (1997).
68. Levine, J. P., Bradley, J. P., Roth, D. A., McCarthy, J. G. & Longaker, M. T. Studies in cranial suture biology: regional dura mater determines overlying suture biology. *Plast. Reconstr. Surg.* **101**, 1441–1447 (1998).
69. Warren, S., Brunet, L., Harland, R. M., Economides, A. & Longaker, M. T. The BMP antagonist noggin regulates cranial suture fusion. *Nature* **422**, 625–629 (2003).
70. Liu, Y. H. *et al.* *Mx2* gene dosage influences the number of proliferative osteogenic cells in growth centers of the developing murine skull: a possible mechanism for *MSX2*-mediated craniosynostosis in humans. *Dev. Biol.* **205**, 260–274 (1999).
71. Satokata, I. *et al.* *Mx2* deficiency in mice causes pleiotropic defects in bone growth and ectodermal organ formation. *Nature Genet.* **24**, 391–395 (2000).
72. Mavrogianis, L. A. *et al.* Haploinsufficiency of the human homeobox gene *ALX4* causes skull ossification defects. *Nature Genet.* **27**, 17–18 (2001).
73. Carter, D. C. & Giori, N. J. in *The Bone-Biomaterial Interface* (ed. Davis, J. E.) 367–379 (University of Toronto Press, Toronto, 1991).
74. Carter, D. R., Beaupré, G. S., Giori, N. J. & Helms, J. A. Mechanobiology of skeletal regeneration. *Clin. Orthopaed. Rel. Res.* **82**, S41–S55 (1998).
75. Probst, A. & Spiegel, H. U. Cellular mechanisms of bone repair. *J. Invest. Surg.* **10**, 77–86 (1997).
76. Colnot, C., Thompson, Z., Miocla, T., Werb, Z. & Helms, J. Altered bone regeneration in the absence of MMP9. *Development* (in the press).
77. Wagner, W. & Harrison, M. R. Fetal operations in the head and neck area: current state. *Head Neck* **24**, 482–490 (2002).
78. Noden, D. M. in *Factors and Mechanisms Influencing Bone Growth* (eds Dixon, A. D. & Sarnat, B. G.) 168–203 (Alan R. Liss, New York, 1982).
79. Lee, S. H., Fu, K. K., Hui, J. N. & Richman, J. M. Noggin and retinoic acid transform the identity of avian facial prominences. *Nature* **414**, 909–912 (2001).
80. Lucas, A. M. & Stettenheim, P. R. *Avian Anatomy: Integument* (United States Department of Agriculture, Washington DC, 1972).
81. Noden, D. M. Origins and patterning of craniofacial mesenchymal tissues. *J. Craniofac. Genet. Dev. Biol.* **2**, 15–31 (1986).

Acknowledgements We gratefully acknowledge helpful discussions with R. Marcucio, C. Kimmel, T. Schilling, M. Bronner-Fraser, D. Noden, M. Longaker, Y. Chai and R. Maxson. This work was supported by NIH grants to J.A.H. and R.A.S.

Developmental regulation of the growth plate

Henry M. Kronenberg

Endocrine Unit, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114-2696, USA
(e-mail: kronenberg.henry@mgh.harvard.edu)

Vertebrates do not look like jellyfish because the bones of their skeletons are levers that allow movement and protect vital organs. Bones come in an enormous variety of shapes and sizes to accomplish these goals, but, with few exceptions, use one process — endochondral bone formation — to generate the skeleton. The past few years have seen an enormous increase in understanding of the signalling pathways and the transcription factors that control endochondral bone development.

Bones are supremely local in their function: they allow adjacent muscles to move them, they move within specialized joint structures, and they protect adjacent organs. But the sizes and shapes of bones need to be carefully coordinated to allow efficient movement of our symmetric bodies. It is not surprising, therefore, that both local paracrine regulators, as well as hormones travelling through the bloodstream, control both bone development and bone remodelling throughout life. The usual suspects — bone morphogenetic proteins (BMPs), Wnts, fibroblast growth factors (FGFs), hedgehog proteins, insulin-like growth factors and retinoids — are essential for normal bone formation. These locally produced factors are joined by systemic factors such as growth hormone, thyroid hormone, oestrogen, androgen, vitamin D and glucocorticoids to control bone growth.

Substantial progress has been made in the past few years in understanding how local signalling molecules, working through key transcription factors, interact and control the growth and differentiation of bones. Here I focus on several intensively studied signalling systems and transcription factors to illustrate the logic of the molecular regulation of bone formation, with an emphasis on how distinct signalling pathways interact to coordinate development.

Endochondral bone formation

Bone formation begins when mesenchymal cells form condensations — clusters of cells that adhere through the expression of adhesion molecules¹ (Fig. 1). In a few areas, most notably the flat bones of the skull, the cells of these condensations differentiate directly into bone-forming osteoblasts. These cells lay down a matrix particularly rich in type I collagen in a process named intramembranous bone formation. But this straightforward process is the exception. In most condensations, the cells become chondrocytes, the primary cell type of cartilage; cells at the border of condensations form a perichondrium. Chondrocytes have a characteristic shape, secrete a matrix rich in type II collagen and the proteoglycan aggrecan, and, more generally, express a characteristic genetic program driven by SOX9 (see below) and other transcription factors.

The cartilage enlarges through chondrocyte proliferation and matrix production (Fig. 1b). Chondrocytes in the centre of the cartilage mould then stop proliferating, enlarge (hypertrophy), and change their genetic program to synthesize type

X collagen (Fig. 1c). The hypertrophic chondrocyte, simply through its size, is the principal engine of bone growth², and is also a master regulatory cell. Hypertrophic chondrocytes direct the mineralization of their surrounding matrix, attract blood vessels through the production of vascular endothelial growth factor and other factors, and attract chondroclasts (closely related or identical to osteoclasts, which are cells of the macrophage lineage that digest matrix). Hypertrophic chondrocytes direct adjacent perichondrial cells to become osteoblasts; these secrete a characteristic matrix, forming a bone collar (Fig. 1d). Hypertrophic chondrocytes then undergo apoptotic cell death. The cartilage matrix left behind provides a scaffold for osteoblasts that invade the cartilage mould along with blood vessels and lay down a true bone matrix within it (primary spongiosa; Fig. 1e).

While hypertrophic chondrocytes perform these multiple tasks in the centre of the cartilage mould, the mould enlarges further through continued proliferation of chondrocytes (Fig. 1f). As the bone enlarges, haematopoietic stem cells interact with the stroma to establish the main site for haematopoiesis in post-natal life. To varying extents, depending on the identity of the particular bone, a portion of these chondrocytes assume a flattened, discoid shape and form columns like stacks of coins with a clear orientation that directs the lengthening of the bone along one axis more than another. Bone lengthening, particularly rapid during fetal life, is driven primarily by the rate of production of hypertrophic chondrocytes from these proliferating chondrocytes. As bones enlarge further, so-called secondary ossification centres are established when chondrocytes in characteristic, new locations stop proliferating, hypertrophy, and attract vascular invasion along with osteoblasts. In the long bones of the limb, growth chondrocytes continue to proliferate between regions of bone of the primary and secondary ossification centres. This cartilage is then called the growth plate, as it forms a distinct plate of cells between the bone of the secondary ossification centre and the primary spongiosa. At the top of the growth plate, round chondrocytes no longer proliferate rapidly and are called resting or reserve chondrocytes, and probably serve as precursors for the flat proliferating columnar chondrocytes (Fig. 1g). In humans, growth plates disappear at the time of adolescence after a burst of pubertal activity.

Signalling pathways that control bone growth

Ihh/PTHrP signalling

Indian hedgehog (Ihh) is a master regulator of bone development, coordinating chondrocyte proliferation, chondrocyte

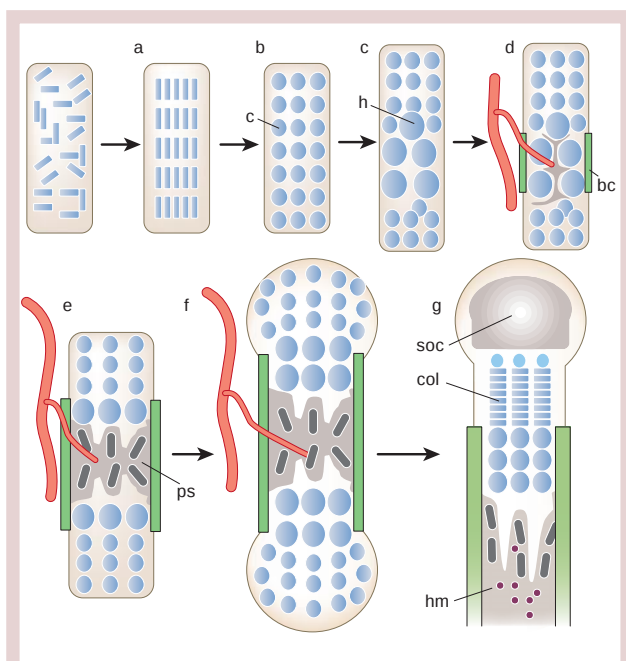


Figure 1 Endochondral bone formation. **a**, Mesenchymal cells condense. **b**, Cells of condensations become chondrocytes (**c**). **c**, Chondrocytes at the centre of condensation stop proliferating and become hypertrophic (**h**). **d**, Perichondrial cells adjacent to hypertrophic chondrocytes become osteoblasts, forming bone collar (**bc**). Hypertrophic chondrocytes direct the formation of mineralized matrix, attract blood vessels, and undergo apoptosis. **e**, Osteoblasts of primary spongiosa accompany vascular invasion, forming the primary spongiosa (**ps**). **f**, Chondrocytes continue to proliferate, lengthening the bone. Osteoblasts of primary spongiosa are precursors of eventual trabecular bone; osteoblasts of bone collar become cortical bone. **g**, At the end of the bone, the secondary ossification centre (**soc**) forms through cycles of chondrocyte hypertrophy, vascular invasion and osteoblast activity. The growth plate below the secondary centre of ossification forms orderly columns of proliferating chondrocytes (**col**). Haematopoietic marrow (**hm**) expands in marrow space along with stromal cells.

differentiation and osteoblast differentiation. *Ihh* is a member of the hedgehog family of secreted ligands, and is thus closely related to Sonic hedgehog, one of the main regulators of limb outgrowth (see review in this issue by Mariani and Martin, page 319). During endochondral bone development, *Ihh* is synthesized by chondrocytes leaving the proliferative pool (prehypertrophic chondrocytes) and by early hypertrophic chondrocytes. *Ihh* binds to its receptor Patched-1 (*Ptc-1*); this binding leads to activation of Smoothened (*Smo*), a membrane protein required for the cellular actions of *Ihh*. Active *Smo* then triggers a cascade that leads to gene activation.

Because *Ihh* action increases expression of *Ptc-1*, changes in levels of *Ptc-1* messenger RNA offer a convenient assay for evidence of *Ihh* action at the cellular level. *Ihh*^{−/−} mice have normal bones at the condensation stage, but subsequently develop pronounced abnormalities of bone growth³. All cartilage elements are small because of a marked decrease in chondrocyte proliferation. The proliferative effect of *Ihh* is likely to be a direct action on chondrocytes, because the expression of *Ptc-1* and of other general targets of hedgehog signalling is stimulated by *Ihh* in chondrocytes. Furthermore, cartilage-specific knockout of *Smo* leads to decreased proliferation of chondrocytes, and either chondrocyte-specific transgenic expression of *Ihh* or of a constitutively active form of *Smo* increases chondrocyte proliferation⁴.

A second abnormality in the bones of *Ihh*^{−/−} mice is an increase in the fraction of chondrocytes that are post-mitotic, hypertrophic chondrocytes. This abnormality results from chondrocytes leaving the pool of proliferating chondrocytes prematurely. They do so

because the cartilage in *Ihh*^{−/−} mice fails to synthesize parathyroid hormone-related protein (PTHrP). PTHrP is a protein secreted during fetal life by perichondrial cells at the ends of cartilage moulds and by early proliferative chondrocytes. PTHrP acts on the same G-protein-coupled receptor used by parathyroid hormone, the calcium-regulating hormone. These PTH/PTHrP receptors (PPRs) are expressed at low levels by proliferating chondrocytes and at high levels by prehypertrophic/early hypertrophic chondrocytes.

PTHrP acts primarily to keep proliferating chondrocytes in the proliferative pool. In *PTHrP*^{−/−} or *PPR*^{−/−} mice, chondrocytes become hypertrophic close to the ends of bones^{5,6}. Conversely overexpression of PTHrP in chondrocytes⁷ or expression of a constitutively active PPR⁸ delays the appearance of hypertrophic chondrocytes. As predicted by this model, expression of a constitutively active PPR in chondrocytes of *Ihh*^{−/−} mice reverses the early hypertrophy in these mice (although it has no effect on the defect in chondrocyte proliferation)⁹. The ability of *Ihh* to stimulate the production of PTHrP and thereby to delay chondrocyte hypertrophy was first shown through viral overexpression of Sonic hedgehog (as an *Ihh* surrogate) in chick limbs and through addition of Hedgehog protein to mouse limbs *in vitro*¹⁰. Hedgehog protein blocked hypertrophy of mouse chondrocytes from wild-type limbs but had no effect on chondrocyte differentiation when added to limbs from *PTHrP*^{−/−} or *PPR*^{−/−} mice. These studies demonstrate the obligatory role for PTHrP signalling in mediating the action of *Ihh* to delay hypertrophy.

The interactions of *Ihh* and PTHrP led to the hypothesis that these two paracrine factors together control the decision of chondrocytes to leave the proliferative pool through a feedback loop (Fig. 2). PTHrP, secreted from cells near the ends of bone, acts on its receptor on proliferating chondrocytes to keep them proliferating. When chondrocytes are no longer sufficiently stimulated by PTHrP, they stop proliferating and synthesize *Ihh*. *Ihh* can then, by mechanisms that are still unknown, stimulate the production of PTHrP at the ends of bone. Experiments using chimaeric mice^{11,12} have shown that this feedback loop actually functions in intact bone. Embryonic stem cells missing the gene encoding PPR were inserted into wild-type blastocysts and the phenotypes of the resultant chimaeric bones were analysed. As expected from the phenotype of the *PPR*^{−/−} mice, the *PPR*^{−/−} chondrocytes differentiate into hypertrophic chondrocytes close to the ends of the chimaeric bones. The *Ihh* formed by these abnormal chondrocytes is, therefore, synthesized much closer to the ends of the bone than normal. This leads to an upregulation of PTHrP expression and a consequent lengthening of the columns of normal proliferative chondrocytes. If the embryonic stem cells are missing both the *PPR* and the *Ihh* genes, then premature hypertrophy of the mutant chondrocytes still occurs, but the *PTHrP* mRNA levels are not changed and the chondrocyte columns formed by the wild-type cells are of normal length.

These studies demonstrate the physiological importance of the interactions of PTHrP and *Ihh* to determine the lengths of proliferative columns in individual bones. These interactions probably also serve to assure optimum coordination between differentiation patterns of individual cartilage columns within a growing bone. In the absence of PTHrP, the border between proliferating and hypertrophic cells is ragged, with some mixing of proliferating and hypertrophic cells. In the wild-type bone, the sharpness of the transition between proliferation and hypertrophy is probably increased by local feedback between *Ihh* and PTHrP production.

A third striking abnormality of *Ihh*^{−/−} mice is the absence of osteoblasts in either the primary spongiosa or the bone collar of bones formed by endochondral development³. Normally, the first perichondrial cells to convert to osteoblasts are adjacent to the *Ihh*-producing prehypertrophic/hypertrophic cells of the growth cartilage. Because these cells normally express *Ptc-1* in response to *Ihh*, the effect of *Ihh* to convert perichondrial cells to osteoblasts is probably a direct one. This hypothesis is further supported by the results of the chimaeric experiments in which *PPR*^{−/−} chondrocytes

mix with wild-type chondrocytes^{11,12}. Ectopic osteoblast formation occurs adjacent to the ectopically positioned hypertrophic chondrocytes. No such ectopic bone forms adjacent to the ectopically positioned *PPR*^{-/-}; *Ihh*^{-/-} hypertrophic chondrocytes in parallel experiments. Thus, *Ihh* controls the differentiation of osteoblasts and designates the location at which this differentiation occurs. Although *Ihh* is absolutely required for osteoblast formation in endochondral bone formation, *Ihh* is not required for formation of osteoblasts in the bones of the skull that form without a cartilage mould, although growth of intramembranous bone of the skull is apparently slowed in *Ihh*^{-/-} mice³.

Thus, *Ihh* is a master regulator of both chondrocyte and osteoblast differentiation during endochondral bone formation. *Ihh* stimulates chondrocyte proliferation directly and, through stimulation of PTHrP synthesis, determines the distance from the end of the bone at which chondrocytes stop proliferating and undergo hypertrophic differentiation. This is also the site at which *Ihh* stimulates formation of the bone collar.

Fibroblast growth factor signalling

Recent genetic studies have shown that FGF signalling crucially regulates chondrocyte proliferation and differentiation. Many of the 22 distinct FGF genes and four FGF receptor genes are expressed at every stage of endochondral bone formation¹³. The probable redundancy of ligand utilization combined with the multiple early effects of FGF receptors in development of bone and other tissues have made genetic analysis of the roles of FGF signalling during bone development a particular challenge. In the earliest stages of endochondral bone development, FGF receptor-2 (FGFR2) is expressed in condensing mesenchyme. Multiple FGFs are expressed in condensations and surrounding mesenchyme, although their roles are not established. One possible role of FGF signalling at this early stage is suggested by the finding that FGFs can stimulate SOX9 expression (see below) in a cultured mesenchymal cell line¹⁴.

As chondrocytes form, proliferating chondrocytes express FGFR3 and prehypertrophic/hypertrophic chondrocytes express FGFR1. Perichondrial cells express FGFR2 (Fig. 3). Each of these receptors has distinct and important roles in bone development; FGFR3 is best understood. Knockout of *Fgfr3* in mice leads to an increased rate of proliferation of chondrocytes and an expansion of the length of chondrocyte columns^{15,16}. Activating point mutations in *Fgfr3* that are found in human chondrodystrophies decrease the rate of proliferation of chondrocytes and lead to shortened, disorganized columns in transgenic mice¹⁷.

Thus, FGF signalling through FGFR3 inhibits proliferation; this inhibition is at least partly through activation of the Janus kinase-signal transducer and activator of transcription-1 (JAK-STAT1) pathway¹⁸. The ligands responsible for this signalling are not fully defined, although the recent observation that knockout of *Fgf18* leads to an increase in chondrocyte proliferation that closely resembles the effect of *Fgfr3* knockout argues strongly that FGF18 is one such relevant ligand^{19,20}. The phenotype of the *Fgf18*-knockout mouse is more severe than that of the *Fgfr3* knockout, in that ossification is delayed in *Fgf18*^{-/-} mice. This delay may reflect lack of activation of *Fgfr1* in hypertrophic chondrocytes and perhaps of *Fgfr2* and -1 in perichondrium (Fig. 3). Studies of bone explants *in vitro* demonstrate that FGF signalling, independent of effects on *Ihh* and PTHrP, accelerates terminal differentiation of hypertrophic chondrocytes²¹.

From the previous discussion of the *Ihh*/PTHrP signalling system, one might not expect that simply increasing or decreasing the rate of proliferation of chondrocytes through changes in FGF signalling would also change the lengths of columns of proliferating chondrocytes. No matter what the rate, one might expect that *Ihh* would change the expression of PTHrP until columns reached their characteristic length. However, knockout of the *Fgfr3* gene increases *Ihh* expression and activation of FGFR3 decreases *Ihh* expression^{13,17}. Studies *in vitro*, in which levels of *Ihh* and PTHrP can be regulated

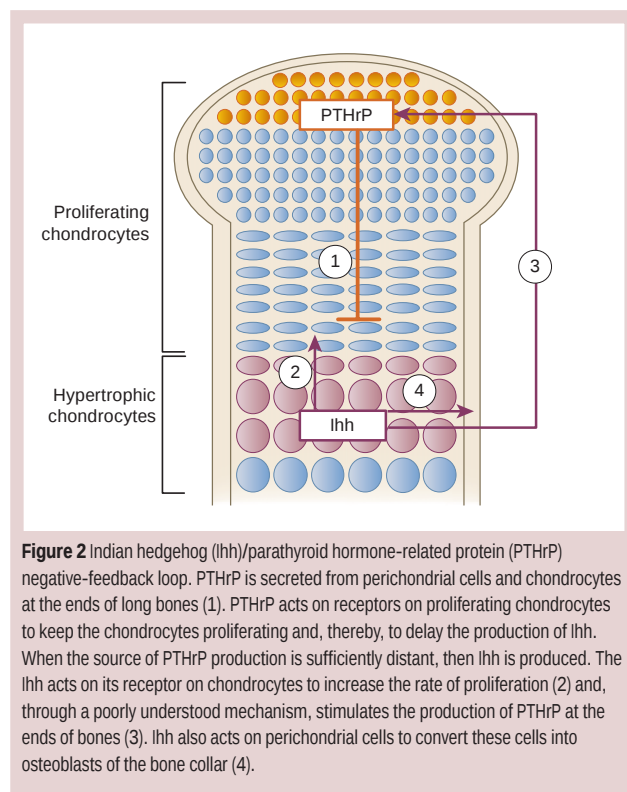


Figure 2 Indian hedgehog (*Ihh*)/parathyroid hormone-related protein (PTHrP) negative-feedback loop. PTHrP is secreted from perichondrial cells and chondrocytes at the ends of long bones (1). PTHrP acts on receptors on proliferating chondrocytes to keep the chondrocytes proliferating and, thereby, to delay the production of *Ihh*. When the source of PTHrP production is sufficiently distant, then *Ihh* is produced. The *Ihh* acts on its receptor on chondrocytes to increase the rate of proliferation (2) and, through a poorly understood mechanism, stimulates the production of PTHrP at the ends of bones (3). *Ihh* also acts on perichondrial cells to convert these cells into osteoblasts of the bone collar (4).

independently, support the idea that part of the effects of FGF signalling is mediated by suppression of *Ihh* expression²¹. Thus, FGF signalling shortens proliferative columns both by decreasing chondrocyte proliferation directly and by suppressing *Ihh* expression.

Bone morphogenetic protein signalling

BMPs, also called growth and differentiation factors (GDFs), have multiple roles during bone formation. As with the FGF system, the large number of ligands and receptors, as well as the vital roles of many components of this network during early development, has precluded a straightforward use of knockout mice to characterize the roles of BMPs in bone development. Nevertheless, it is clear that BMPs/GDFs have essential roles at every stage of endochondral bone development.

BMPs and GDFs are members of the TGF- β family of paracrine factors that activate heterodimeric receptors with serine/threonine kinase activity²². They were discovered because of their remarkable ability to induce endochondral bone formation when injected subcutaneously in mice. The type 1 receptor BMPR1B is expressed in cartilage condensations and the type 1 receptor BMPR1A is expressed broadly in embryonic mesenchyme. An important role for BMP signalling in formation of mesenchymal condensations is supported by the suppression of formation of condensations when the BMP antagonist, Noggin, is expressed in early chick limbs²³ and by the enlarged cartilage primordial in *Noggin*^{-/-} mice²⁴. Several condensations are either abnormal or absent in short ear mice, which harbour various inactivating mutations of the *BMP5* gene²⁵. Mice missing the gene encoding GDF5 or BMPR1B have abnormalities in digit formation that involve failure of extension of condensations that lead to digit formation²⁶⁻²⁸. Understanding of the full role of BMPs and their receptors during condensation formation, however, will require further studies that generate mice with multiple deficiencies of BMPs, their receptors, and the plethora of BMP signalling modulators specifically in early condensations.

BMPs have multiple important roles during later stages of cartilage development. BMP2, -3, -4, -5 and -7 are expressed in

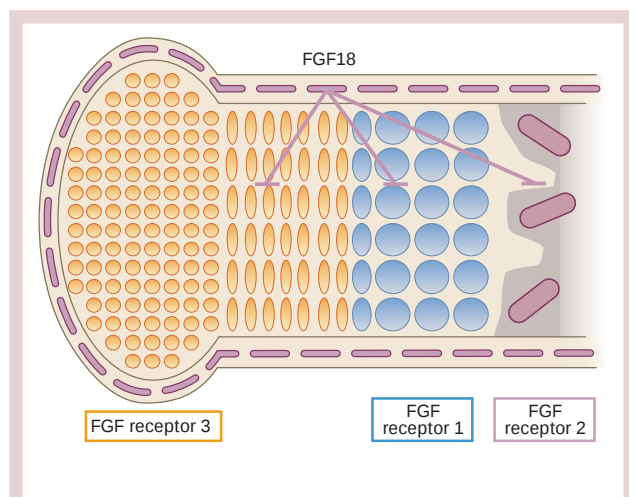


Figure 3 Fibroblast growth factor (FGF) signalling in the growth plate. FGF receptor 3 (FGFR3; yellow) is expressed in proliferating chondrocytes, FGFR1 (light blue) is expressed in prehypertrophic and hypertrophic chondrocytes and perichondrium (not shown), and FGFR2 (purple) is expressed in the perichondrium, periosteum and the primary spongiosa. FGF18 is expressed in the perichondrium and the phenotype of its knockout suggests that it acts on FGFR3 to decrease chondrocyte proliferation and, perhaps, on FGFR1 to delay terminal differentiation of hypertrophic chondrocytes and on FGFR1 and -2 to delay osteoblast development. FGF7, -8 and -17 (not shown) are also expressed in the perichondrium, although knockout of these genes either has no phenotype or leads to early fetal demise, so their function in the growth plate is unknown.

perichondrium, BMP2 and -6 are expressed in hypertrophic chondrocytes, and BMP7 is expressed in proliferating chondrocytes. GDF5, -6 and -7 are expressed at sites of subsequent joint formation. Addition of BMPs to bone explants increases proliferation of chondrocytes and Noggin blocks chondrocyte proliferation^{21,29}. Furthermore, addition of BMP2 delays terminal differentiation of hypertrophic chondrocytes, as indicated by expression of osteopontin, and Noggin hastens terminal hypertrophic differentiation^{21,29}. These actions of BMPs, established using various overexpression systems, need further confirmation with experiments that knock out selective genes.

BMP signalling increases the expression of *Ihh* by prehypertrophic chondrocytes^{21,29}, and so can increase both the proliferation of chondrocytes and the length of proliferating columns of chondrocytes. In both of these actions BMPs oppose the effects of FGF signalling to decrease chondrocyte proliferation and to decrease *Ihh* expression. Because BMPs and FGFs also have opposite effects on terminal hypertrophic chondrocyte differentiation, these pathways can be considered as antagonizing each other at several levels²¹ (Fig. 4). The molecular mechanisms of BMP/FGF antagonism are unknown, although it does not involve simple suppression of ligand expression, as FGF2 increases BMP7 expression and BMP7 increases FGF18 expression.

Transcription factors central to growth plate function SOX9

SOX9 is essential for converting cells of condensations into chondrocytes and acts further at every stage of chondrocyte differentiation. SOX9 is closely related to the Y-chromosome-encoded testis-determining factor SRY (the abbreviation 'SOX' refers to the DNA-binding SRY box found in SOX-family members) and other DNA-binding proteins in the HMG (high mobility group) family. SOX9 is expressed in cells of mesenchymal condensations and in proliferating chondrocytes, but not in hypertrophic chondrocytes. In cultured cells, SOX9 stimulates transcription of a number of cartilage matrix genes, including *Col2a1*, *Col11a2* and *aggrecan*. In

transgenic mice, ectopic expression of *Sox9* leads to expression of *Col2a1* in brain.

The essential role for SOX9 in chondrocyte formation was demonstrated by analysis of chimaeric mice containing descendants of *Sox9*^{-/-} embryonic stem cells in wild-type hosts³⁰. Mesenchymal cells of the *Sox9*^{-/-} lineage could not participate in condensations or subsequent formation of chondrocytes. When a Cre-loxP strategy (a powerful technology that allows target genes to be spliced out of specific cells and tissues) was used to knock *Sox9* out selectively from early limb mesenchyme, no cartilage condensations formed and increased apoptosis in the mesenchyme was noted³¹. Thus, *Sox9* may direct differentiation and thereby allow survival of cells in condensations.

When *Sox9* was deleted from chondrocytes at later stages of development, through the use of Cre driven by a *Col2a1* promoter, the resultant chondrocytes displayed decreased proliferation, decreased expression of matrix genes, and decreased expression of elements of the *Ihh*-PTHrP signalling pathways. These phenotypes closely resemble those of mice missing both *Sox5* and *Sox6*, other Sox-family members that cooperate with *Sox9* in regulating many chondrocyte genes³² and which are not expressed in the absence of *Sox9*³¹. Even heterozygous deletion of *Sox9* in mice leads to cartilage hypoplasia and a perinatal lethal osteochondrodysplasia that resembles campomelic dysplasia, a human syndrome caused by haploinsufficiency of *SOX9*³¹. These mice have fewer chondrocytes than normal in all bone anlage and malformation of subsequent bones. Furthermore, proliferative chondrocytes seem to convert prematurely to hypertrophic chondrocytes, which then mineralize their matrix prematurely. Thus, SOX9 is critical for all phases of the chondrocyte lineage from early condensations to the conversion of proliferating chondrocytes to hypertrophic chondrocytes.

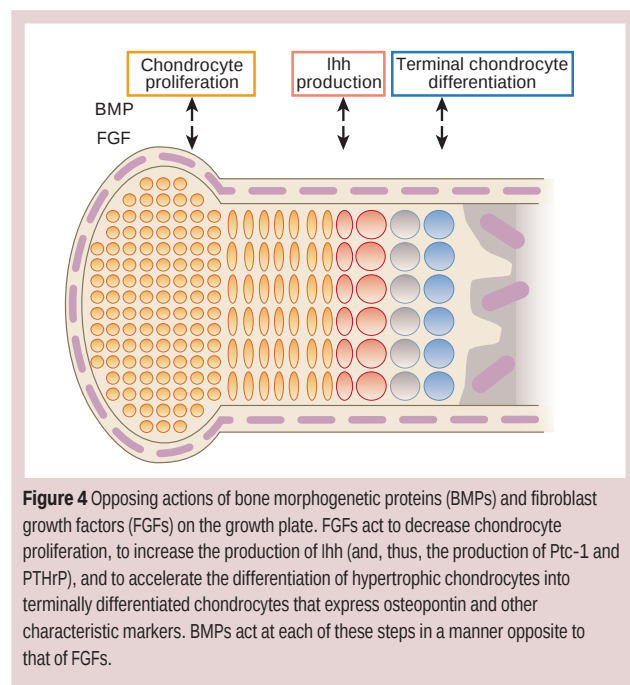
Runx2

Runx2 (previously called Cbfa1) came to prominent attention because mice missing *Runx2* have no osteoblasts^{33,34}. It is also important in the growth plate, as it drives proliferative chondrocytes to differentiate further into hypertrophic chondrocytes. Runx2 is a member of a family of transcription factors that share the DNA-binding domain of the *Drosophila* pair rule gene *runt*. Runx2 is expressed in the late condensation stage of chondrogenesis and then has substantially decreased expression in proliferating chondrocytes, with increased expression again in prehypertrophic and hypertrophic chondrocytes. It is highly expressed in perichondrial cells and in osteoblasts.

Mice missing *Runx2* have no osteoblasts and also exhibit abnormalities of chondrocyte maturation³³⁻³⁵. Most bones either lack or have decreased numbers of hypertrophic chondrocytes, and the hypertrophic chondrocytes that are present fail to mineralize their matrix and have decreased or absent expression of genes such as *osteopontin* and *matrix metalloproteinase 13*, which are normally expressed by late hypertrophic chondrocytes. These abnormalities can be reversed by transgene-driven expression of *Runx2* in chondrocytes^{36,37}. Transgenic expression of *Runx2* in wild-type mice accelerates hypertrophy of normal chondrocytes and even induces hypertrophy and subsequent bone formation in cartilage that normally never undergoes hypertrophy, such as the cartilage of tracheal rings. Transgenic expression of a dominant-negative form of *Runx2* blocks hypertrophy of all chondrocytes³⁷. The hypertrophy that occurs in the absence of *Runx2* may result from the action of other Runx-family members, such as *Runx3*.

Prospects

In this brief overview, the focus on a small number of interacting signalling systems and transcription factors only begins to analyse the number of important signals that determine endochondral bone formation. As noted earlier, growth plate development is affected by growth hormone and insulin-like growth factors, Wnts, growth



factors, cytokines, and nuclear receptors responsive to retinoids, thyroid hormone, oestrogen, androgen, vitamin D and glucocorticoids, among others. Furthermore, even with the best-characterized signalling systems, the crucial target transcription factors that mediate the actions of these signals are poorly understood.

More generally, many important biological phenomena described here have no clear molecular explanation. The determinants of the polarity of proliferative columns — both the orientation of the columns within the bone and the asymmetry associated with hypertrophic differentiation occurring at one characteristic end of columns — are not known. The control of the profound and progressive slowing of bone growth in late fetal and then early post-natal life is not understood. The pathways used to determine the enormous variety of specific bone shapes and sizes that all arise using the common endochondral development process are unknown. The determinants of chondrocyte shape — small and round becoming flat and then hypertrophic — are unknown, but crucial for an understanding of bone lengthening. The coordination of growth plate function with the development of joints, tendons and ligaments is under study, but our understanding remains preliminary. The signals that bring haematopoietic stem cells to bone marrow have not yet been clarified. Fortunately, the increasingly powerful tools now available make all of these questions accessible to investigation. □

doi:10.1038/nature01657

- Hall, B. K. & Miyake, T. All for one and one for all: condensations and the initiation of skeletal development. *BioEssays* **22**, 138–147 (2000).
- Noonan, K. J., Hunziker, E. B., Nessler, J. & Buckwalter, J. A. Changes in cell, matrix compartment, and fibrillar collagen volumes between growth-plate zones. *J. Orthop. Res.* **16**, 500–508 (1998).
- St-Jacques, B., Hammerschmidt, M. & McMahon, A. P. Indian hedgehog signaling regulates proliferation and differentiation of chondrocytes and is essential for bone formation. *Genes Dev.* **13**, 2072–2086 (1999). [Published erratum appears in *Genes Dev.* **13**, 2617 (1999).]
- Long, F., Zhang, X. M., Karp, S., Yang, Y. & McMahon, A. P. Genetic manipulation of hedgehog signaling in the endochondral skeleton reveals a direct role in the regulation of chondrocyte proliferation. *Development* **128**, 5099–5108 (2001).
- Karaplis, A. C. *et al.* Lethal skeletal dysplasia from targeted disruption of the parathyroid hormone-related peptide gene. *Genes Dev.* **8**, 277–289 (1994).

- Lanske, B. *et al.* PTH/PTHrP receptor in early development and Indian hedgehog-regulated bone growth. *Science* **273**, 663–666 (1996).
- Weir, E. C. *et al.* Targeted overexpression of parathyroid hormone-related peptide in chondrocytes causes chondrodysplasia and delayed endochondral bone formation. *Proc. Natl Acad. Sci. USA* **93**, 10240–10245 (1996).
- Schipani, E. *et al.* Targeted expression of constitutively active receptors for parathyroid hormone and parathyroid hormone-related peptide delays endochondral bone formation and rescues mice that lack parathyroid hormone-related peptide. *Proc. Natl Acad. Sci. USA* **94**, 13689–13694 (1997).
- Karp, S. J. *et al.* Indian Hedgehog coordinates endochondral bone growth and morphogenesis via parathyroid hormone related-protein-dependent and -independent. *Development* **127**, 543–548 (2000).
- Vortkamp, A. *et al.* Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* **273**, 613–622 (1996).
- Chung, U. I., Lanske, B., Lee, K., Li, E. & Kronenberg, H. The parathyroid hormone/parathyroid hormone-related peptide receptor coordinates endochondral bone development by directly controlling chondrocyte differentiation. *Proc. Natl Acad. Sci. USA* **95**, 13030–13035 (1998).
- Chung, U. I., Schipani, E., McMahon, A. P. & Kronenberg, H. M. Indian hedgehog couples chondrogenesis to osteogenesis in endochondral bone development. *J. Clin. Invest.* **107**, 295–304 (2001).
- Ornitz, D. M. & Marie, P. J. FGF signaling pathways in endochondral and intramembranous bone development and human genetic disease. *Genes Dev.* **16**, 1446–1465 (2002).
- Murakami, S., Kan, M., McKeehan, W. L. & de Crombrugge, B. Up-regulation of the chondrogenic Sox9 gene by fibroblast growth factors is mediated by the mitogen-activated protein kinase pathway. *Proc. Natl Acad. Sci. USA* **97**, 1113–1118 (2000).
- Deng, C., Wynshaw-Boris, A., Zhou, F., Kuo, A. & Leder, P. Fibroblast growth factor receptor 3 is a negative regulator of bone growth. *Cell* **84**, 911–921 (1996).
- Colvin, J. S., Bohne, B. A., Harding, G. W., McEwen, D. G. & Ornitz, D. M. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nature Genet.* **12**, 390–397 (1996).
- Naski, M. C., Colvin, J. S., Coffin, J. D. & Ornitz, D. M. Repression of hedgehog signaling and BMP4 expression in growth plate cartilage by fibroblast growth factor receptor 3. *Development* **125**, 4977–4988 (1998).
- Sahni, M. *et al.* FGF signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway. *Genes Dev.* **13**, 1361–1366 (1999).
- Liu, Z., Xu, J., Colvin, J. S. & Ornitz, D. M. Coordination of chondrogenesis and osteogenesis by fibroblast growth factor 18. *Genes Dev.* **16**, 859–869 (2002).
- Ohbayashi, N. *et al.* FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis. *Genes Dev.* **16**, 870–879 (2002).
- Minina, E., Kreschel, C., Naski, M. C., Ornitz, D. M. & Vortkamp, A. Interaction of FGF, Ihh/PTHrP, and BMP signaling integrates chondrocyte proliferation and hypertrophic differentiation. *Dev. Cell* **3**, 439–449 (2002).
- Rosen, V. & Wozney, J. M. in *Principles of Bone Biology* (eds Bilezikian, J. P., Raisz, L. G. & Rodan, G. A.) 919–928 (Academic, San Diego, 2002).
- Pizette, S. & Niswander, L. BMPs are required at two steps of limb chondrogenesis: formation of prechondrogenic condensations and their differentiation into chondrocytes. *Dev. Biol.* **219**, 237–249 (2000).
- Brunet, L. J., McMahon, J. A., McMahon, A. P. & Harland, R. M. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science* **280**, 1455–1458 (1998).
- Kingsley, D. M. *et al.* The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGF beta superfamily. *Cell* **71**, 399–410 (1992).
- Storm, E. E. & Kingsley, D. M. GDF5 coordinates bone and joint formation during digit development. *Dev. Biol.* **209**, 11–27 (1999).
- Yi, S. E., Daluiski, A., Pederson, R., Rosen, V. & Lyons, K. M. The type I BMP receptor BMPRII is required for chondrogenesis in the mouse limb. *Development* **127**, 621–630 (2000).
- Baur, S. T., Mai, J. J. & Dymecki, S. M. Combinatorial signaling through BMP receptor IB and GDF5: shaping of the distal mouse limb and the genetics of distal limb diversity. *Development* **127**, 605–619 (2000).
- Minina, E. *et al.* BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation. *Development* **128**, 4523–4534 (2001).
- Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. & de Crombrugge, B. Sox9 is required for cartilage formation. *Nature Genet.* **22**, 85–89 (1999).
- Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & de Crombrugge, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* **16**, 2813–2828 (2002).
- Smits, P. *et al.* The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev. Cell* **1**, 277–290 (2001).
- Otto, F. *et al.* *Cbfa1*, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89**, 765–771 (1997).
- Komori, T. *et al.* Targeted disruption of *Cbfa1* results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* **89**, 755–764 (1997).
- Inada, M. *et al.* Maturational disturbance of chondrocytes in *Cbfa1*-deficient mice. *Dev. Dyn.* **214**, 279–290 (1999).
- Takeda, S., Bonnamy, J. P., Owen, M. J., Ducey, P. & Karsenty, G. Continuous expression of *Cbfa1* in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues *Cbfa1*-deficient mice. *Genes Dev.* **15**, 467–481 (2001).
- Ueta, C. *et al.* Skeletal malformations caused by overexpression of *Cbfa1* or its dominant negative form in chondrocytes. *J. Cell Biol.* **153**, 87–100 (2001).

Acknowledgements I thank A. McMahon for a careful review of this manuscript.

Osteoclast differentiation and activation

William J. Boyle*, W. Scott Simonet† & David L. Lacey†

*Protein Pathways, Inc., Woodland Hills, California 91367, and †Amgen, Inc., Thousand Oaks, California 91320, USA
(e-mail: bboyle@proteinpathways.com)

Osteoclasts are specialized cells derived from the monocyte/macrophage haematopoietic lineage that develop and adhere to bone matrix, then secrete acid and lytic enzymes that degrade it in a specialized, extracellular compartment. Discovery of the RANK signalling pathway in the osteoclast has provided insight into the mechanisms of osteoclastogenesis and activation of bone resorption, and how hormonal signals impact bone structure and mass. Further study of this pathway is providing the molecular basis for developing therapeutics to treat osteoporosis and other diseases of bone loss.

In life, bone is a rigid yet dynamic organ that is continuously moulded, shaped and repaired. Bone microstructure is patterned to provide maximal strength with minimal mass, as determined by the physiological needs of the organism. How are bone structure and function maintained, and how are changes in bone metabolism induced? Once formed, bone undergoes a process termed remodelling that involves break down (resorption) and build-up (synthesis) of bone; this occurs in micro scale throughout the skeleton. Bone remodelling is the predominant metabolic process regulating bone structure and function during adult life, with the key participant being the osteoclast^{1,2}. Imbalances of remodelling can result in gross perturbations in skeletal structure and function, and potentially to morbidity and shortening of lifespan.

Most adult skeletal diseases are due to excess osteoclastic activity, leading to an imbalance in bone remodelling which favours resorption³. Such diseases would include osteoporosis, periodontal disease, rheumatoid arthritis, multiple myeloma and metastatic cancers. For individuals with osteoporosis, bone fractures represent life-threatening events, and today there are in excess of 70 million people worldwide at risk. Recent breakthroughs in our understanding of osteoclast differentiation and activation have come from the analysis of a family of biologically related tumour necrosis factor (TNF) receptor (TNFR)/TNF-like proteins: osteoprotegerin (OPG), receptor activator of nuclear factor (NF)- κ B (RANK) and RANK ligand (RANKL), which together regulate osteoclast function⁴. The study of this pathway is providing a deeper understanding of how diverse physiological and pathophysiological signals exert their effects on the RANK signalling pathway to induce osteoclastogenesis, bone resorption and skeletal remodelling, and so control bone mass.

Osteoclastogenesis

The osteoclast is a tissue-specific macrophage polykaryon created by the differentiation of monocyte/macrophage precursor cells at or near the bone surface (Fig. 1). A significant breakthrough in our understanding of osteoclastogenesis occurred when murine systems using co-cultures of bone marrow or spleen cells and stromal cells yielded osteoclasts⁵. The close contact between stromal and bone marrow cell types was essential for osteoclastogenesis, and suggested that

stromal-derived factors stimulate this process. It is now known that this system allowed for production of two haematopoietic factors that are both necessary and sufficient for osteoclastogenesis, the TNF-related cytokine RANKL and the polypeptide growth factor CSF-1 (for colony-stimulating factor-1)^{6,7}, and for the subsequent activation of RANK on the surface of haematopoietic precursor cells^{8,9}. Together, CSF-1 and RANKL are required to induce expression of genes that typify the osteoclast lineage, including those encoding tartrate-resistant acid phosphatase (TRAP), cathepsin K (CATK), calcitonin receptor and the β_3 -integrin⁷, leading to the development of mature osteoclasts.

The mature, multinucleated osteoclast is activated by signals, which leads to initiation of bone remodelling (Fig. 2). The osteoclast cell body is polarized, and in response to activation of RANK by its ligand¹⁰, it undergoes internal structural changes that prepare it to resorb bone, such as the rearrangements of the actin cytoskeleton and formation of a tight junction between the bone surface and basal membrane to form a sealed compartment. This external 'vacuole' is then acidified by the export of hydrogen ions generated by the ATP6i complex¹¹. Secretion continues with the export of the lytic enzymes TRAP and pro-CATK into a resorption pit (Howship's lacunae). Through this process the osteoclast erodes the underlying bone. Degradation products (collagen fragments and solubilized calcium and phosphate) are processed within the osteoclast and released into the circulation. RANKL can activate mature osteoclasts in a dose-dependent manner *in vitro*, and can lead rapidly to the resorption of bone *in vivo* by activating pre-existing osteoclasts^{10,12}. The survival of the mature osteoclast, and its participation in successive rounds of bone resorption, is regulated in part by hormones and cytokines¹³. RANKL and interleukin (IL)-1 increase the survival time of the mature osteoclast *in vitro* and *in vivo*, and this may be due to the ability of these factors to induce NF- κ B activity^{14,15}.

At least 24 genes or loci have been shown to positively and negatively regulate osteoclastogenesis and osteoclast activation (Fig. 1), based on naturally occurring mutations or targeted knockout mutations in rodents and humans (see refs 16, 17 for review). Disruption of these genes blocks the development and/or function of the mature osteoclast, resulting in abnormally high levels of mineralized bone and cartilage, a condition called osteopetrosis. The gene products for some of these loci have yet to be identified, although

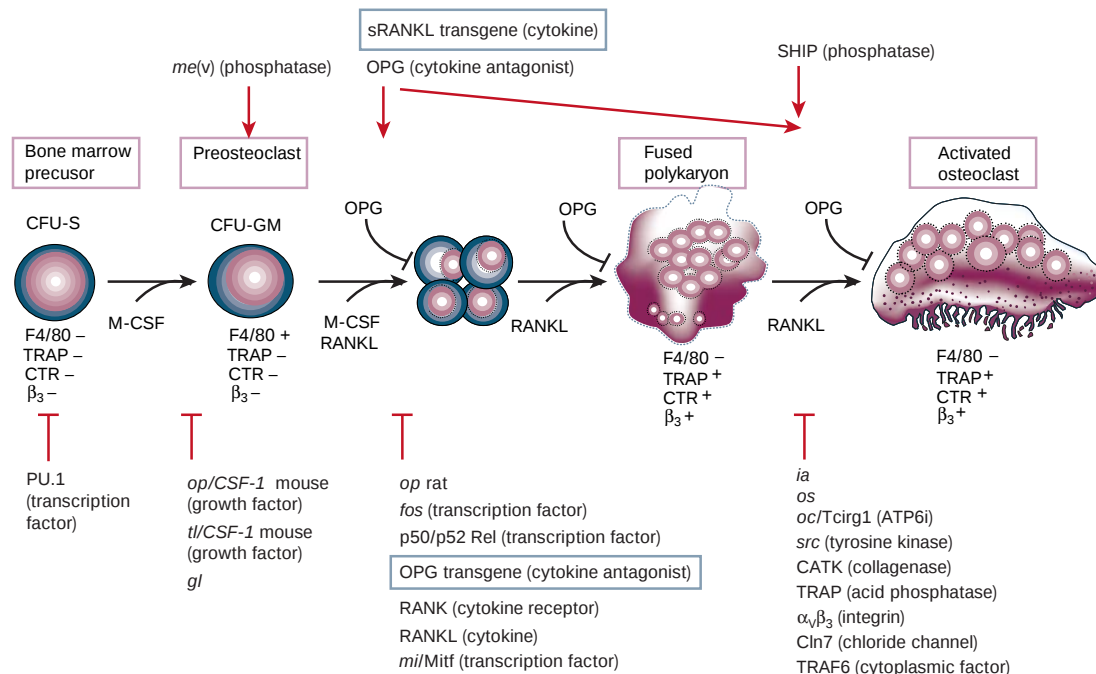


Figure 1 Osteoclastogenesis. Development schema of haematopoietic precursor cell differentiation into mature osteoclasts, which are fused polykaryons arising from multiple (10–20) individual cells. Maturation occurs on bone from peripheral blood-borne mononuclear cells with traits of the macrophage lineage shown below. M-CSF (CSF-1) and RANKL are essential for osteoclastogenesis, and their action during lineage allocation and maturation is shown. OPG can bind and neutralize RANKL, and can negatively regulate both osteoclastogenesis and activation of mature osteoclasts.

Shown below are the single-gene mutations that block osteoclastogenesis and activation. Those indicated in *italic font* are naturally occurring mutations in rodents and humans, whereas the others are the result of targeted mutagenesis to generate null alleles. Shown above are the single-gene mutant alleles that increase osteoclastogenesis and/or activation and survival and result in osteoporosis. Note that all of these mutants represent null mutations with the exception of the OPG²² and sRANKL⁷⁷ transgenic mouse overexpression models (in blue-outlined boxes).

the functions of the *microphthalmic* (*mi*; transcription factor), *toothless* (*tl*; growth factor) and *osteosclerotic* (*oc*; ATPase) gene products have recently been discovered^{17,18}. Targeted disruption of some genes, such as those encoding SHIP (Src homology 2 domain-containing inositol-5-phosphatase) and osteoprotegerin, cause increased osteoclastogenesis and/or activation *in vitro* and *in vivo*, resulting in osteopenia^{19,20}. As a class of mutations, these genes exert their effects at various stages of osteoclast development and activation (Fig. 1). Some genes act during the formation and/or survival of the osteoclast precursor cell (*PU.1* and *op/CSF-1*), whereas other genes mediate either the ability of the precursor cell to undergo differentiation (*RANK*, *p50/p52 rel* and *fos*) or the adherence and lytic function of the mature osteoclast (*src*, *oc/Tcigr* and *CATK*). Continued analysis of the unknown loci in this group of genes is likely to further define important control mechanisms within the osteoclast and the osteoclast regulatory networks²¹.

The RANKL/RANK/OPG regulatory axis

The bone marrow and stromal cell co-culture system used to generate osteoclasts *in vitro* was also useful for testing the effects of soluble factors on osteoclast formation. A breakthrough in our understanding of how osteoclastogenesis is regulated was the identification of OPG, a soluble protein that blocked osteoclast formation *in vitro* and bone resorption *in vivo*. OPG is a secreted TNFR-related protein that regulates bone density and bone mass in animals^{22,23}, and upon systemic administration can block pathologic bone resorption in various animal models^{22,24}. Because the TNFR motifs of OPG were found to be essential for its biological activity, it was not surprising that the TNF-related surface protein RANKL was identified as a key cytokine that regulated osteoclastogenesis and bone resorp-

tion^{6,7,25,26}. RANKL was subsequently shown to bind and activate the TNFR-related protein RANK, a transmembrane signalling receptor²⁶. RANK expression on haematopoietic precursor cells is required in the mouse for osteoclast differentiation and activation, and for the resorption of bone and regulation of calcium homeostasis by calcitropic hormones^{27,28}. Mice deficient in RANK or RANKL are phenocopies of one another, indicating the essential role of this receptor–ligand pair in bone remodelling^{27,29}.

The RANKL polypeptide is a type II transmembrane protein found on the surface of expressing cells as a proteolytically released soluble form^{7,25,26}. Hormones and factors that stimulate bone resorption *in vivo* induce the expression of RANKL on osteogenic stromal cells^{30,31}. RANKL expression by osteoblasts coordinates bone remodelling by stimulating bone resorption by local osteoclasts, which in turn stimulate bone synthesis by closely adjacent osteoblasts by a process called ‘coupling’³². OPG therefore acts as a decoy receptor by blocking RANKL binding to its cellular receptor RANK. OPG is also produced by osteoblasts in response to anabolic agents such as oestrogens and transforming growth factor-β (TGF-β)-related bone morphogenic proteins (BMPs)^{32,33}. OPG overexpression blocks osteoclast production, which leads to osteopetrosis in mice, whereas OPG deletion results in enhanced remodelling of bone and osteoporosis^{21,22}. Expression of RANKL and OPG is therefore coordinated to regulate bone resorption and density positively and negatively by controlling the activation state of RANK on osteoclasts.

RANK signalling network in osteoclasts

Activation of RANK by its ligand leads to the expression of osteoclast-specific genes during differentiation, the activation of resorption by mature osteoclasts, and their survival and participation in new

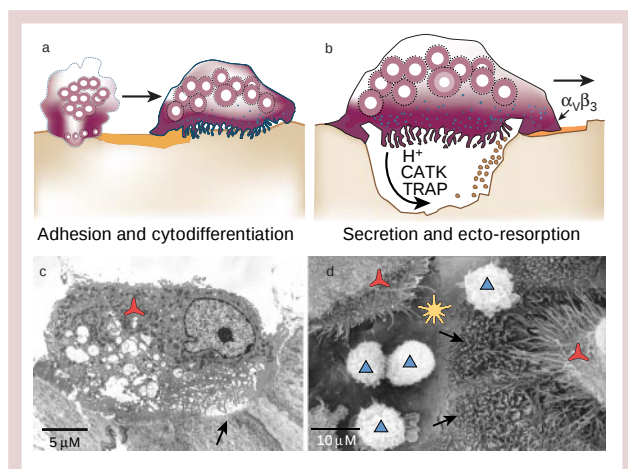


Figure 2 Activation of bone resorption. **a**, Multinucleated polykaryons are recruited by the action of CSF-1 and RANKL, which then adhere to bone and undergo cytodifferentiation into a mature osteoclast. **b**, RANKL stimulates osteoclast activation by inducing secretion of protons and lytic enzymes into a sealed resorption vacuole formed between the basal surface of the osteoclast and the bone surface. Acidification of this compartment by secretion of protons leads to the activation of TRAP and CATK, which are the two main enzymes responsible for the degradation of bone mineral and collagen matrices. **c**, Transmission electron micrograph of an activated mouse osteoclast with a visible ruffled border in a resorption lacuna on the periosteal femoral cortical bone surface. Red propeller, osteoclast; black arrow, a resorption pit. **d**, Scanning electron micrograph of human osteoclasts generated *in vitro* on cortical bone slices from CSF-1- and RANKL-treated peripheral blood mononuclear cells. Red propellers, osteoclasts; black arrows, a resorption pit where the normally smooth lamellar bone surface has been resorbed to expose collagen bundles; yellow star, non-resorbed bone surface; blue triangles, mononuclear cells (potential osteoclast precursors).

rounds of bone degradation at neighbouring sites (Fig. 3). RANK signalling is mediated by cytoplasmic factors that activate downstream signalling pathways that control these various functions. At least five distinct signalling cascades mediated by protein kinases are induced during osteoclastogenesis and activation—inhibitor of NF- κ B kinase (IKK), c-Jun N-terminal kinase (JNK), p38, extracellular signal-regulated kinase (ERK) and Src pathways (Fig. 3). The key preliminary step in RANK signalling is the binding of TNFR-associated cytoplasmic factors, or TRAFs, to specific domains within the cytoplasmic domain of RANK^{34,35}. TRAF2, -5 and -6 have all been shown to bind to RANK, and their individual docking sites have been mapped, yet only TRAF6 mutations result in osteopetrosis due to a loss of osteoclast activity^{36,37}. RANK mutants that specifically lack TRAF6-binding sites are unable to restore osteoclastogenic potential in RANK^{-/-}-derived haematopoietic precursors³⁸. TRAF6 binding to RANK has been modelled, and involves a new structure–function relationship between TRAFs and TNFR-related proteins³⁹.

TRAF6 acts as a key adaptor to assemble signalling proteins that direct osteoclast-specific gene expression leading to differentiation and activation. The two most closely studied pathways are the activation of transcription factors NF- κ B and activator protein-1 (AP-1), whose activities are rapidly induced following ligand binding. Targeted mutagenesis of the p50/p52 component of NF- κ B^{40,41}, as well as the cFos component of AP-1⁴², results in osteopetrosis owing to a block in osteoclastogenesis. Activation of these transcription factors can be induced by signalling cascades mediated by protein kinases, including IKK1/2 (NF- κ B) and JNK1 (AP-1)^{43,44}. Recently, the mitogen-activated protein kinase (MAPK)-related TGF- β -inducible kinase TAK1, along with the TRAF-binding adapter protein TAB2, have been detected in activated receptor complexes^{45,46}. Dominant-interfering mutant forms of TAK1 inhibit

RANKL-mediated activation of both IKK1/2 and JNK1, suggesting that TAK1 is important in activation of NF- κ B and AP-1⁴⁷. In addition, the MAPK-related kinase MKK7 is required for JNK activation in these cells. It is not yet clear if TAK1 acts directly on IKK1/2 or MKK7, or if other kinases mediate these events (Fig. 3).

The stress-activated protein kinase p38 is also involved in mediating key signals induced by RANK, and is apparently activated via phosphorylation by MKK6^{48,49}. Stimulation of p38 results in the downstream activation of the transcriptional regulator mi/Mitf, which controls the expression of the genes encoding TRAP and CATK⁵⁰. mi/Mitf, CATK and TRAP are all required by the mature osteoclast (Fig. 1), indicating the importance of p38 signal transduction. The ERK-1 kinase is also activated by RANK signalling, and seems to be regulated upstream by activation of MEK1^{51,52}. The small molecule ERK inhibitors PD98059 and U0126 potentiate RANKL-induced osteoclast differentiation, suggesting that the ERK pathway is involved in negative regulation of osteoclastogenesis.

The Src protein, which is required for osteoclast activation, has also been shown to bind to TRAF6 and to allow RANK-mediated signalling to proceed through the lipid kinase phosphatidylinositol 3-OH kinase (PI(3)K) and the serine/threonine protein kinase AKT⁵³. Both PI(3)K and AKT are known to act downstream of Src to induce cell survival, cytoskeletal rearrangements and motility. The lipid kinase SHIP negatively regulates PI(3)K signalling, and SHIP-deficient mice have osteoporosis¹⁹. The immunosuppressant rapamycin, a small molecule inhibitor of mTOR (FRAP), a downstream substrate of AKT, enhances osteoclastogenesis *in vitro*⁵⁴.

Signalling from RANK culminates in the change in gene expression patterns that characterize the active osteoclast. Several groups have used gene expression arrays to map these changes in cell culture systems^{55–57}. A consistent finding among these groups is the considerable upregulation of nuclear factor of activated T cells (NFAT)-2, a calcineurin- and calcium-regulated transcription factor. Inhibiting NFAT2 activity using dominant negative alleles blocks osteoclastogenesis, whereas overexpression of the wild-type protein stimulated osteoclast development from embryonic stem cells in a RANKL-independent manner⁵⁷. Similarly, the Myc nucleoprotein factor is induced during differentiation, and its activity is required for *in vitro* osteoclastogenesis⁵⁸. It is not clear which signalling pathway(s) lead to activation of NFAT2 and Myc, although the activation of NFAT2 is blocked by the calcineurin inhibitors cyclosporin A and FK506⁵⁷.

Modulation of RANK-induced osteoclastogenesis

There are several levels of control of the RANK signalling pathway that enhance or dampen osteoclastogenesis and activation driven by RANKL. Activation of osteoclast surface receptors for IL-1, c-Fms, TNF- α , PGE2 and TGF- β potentiate osteoclastogenesis *in vitro*, and can stimulate bone resorption *in vivo*. IL1-R and TNFR1 both signal through TRAF6, and activation of these receptors could have a synergistic effect on RANK-mediated TRAF6 activation⁵⁹. The activation of c-Fms and TGF- β has been reported to upregulate components of the pathway, including the levels of RANK on the cell surface, thereby impacting the potency of RANKL in this system^{60–62}.

The RANK signalling pathway is negatively controlled by OPG *in vitro* and *in vivo*^{22,23}. In addition to this, chemical inhibitors of MEK1 and mTOR lead to increased osteoclastogenesis *in vitro*, suggesting that activation of the Erk and Src pathways can also negatively regulate osteoclastogenesis^{52,54}. There is also evidence for feedback mechanisms that switch off the RANK signalling pathway once it is activated. Induction of osteoclastogenesis by RANKL leads to the induction of interferon- β , which is secreted and acts in an autocrine fashion to downregulate the expression of c-Fos, a critical factor involved in osteoclast development^{63,64}. Interferon- γ (INF- γ) also has a negative effect on this pathway. Binding of INF- γ to its receptor leads to degradation of TRAF6, and can have an inhibitory effect on osteoclastogenesis *in vitro*⁶⁵. This result is surprising, as INF- γ is an approved therapeutic for the treatment of osteopetrosis, where its

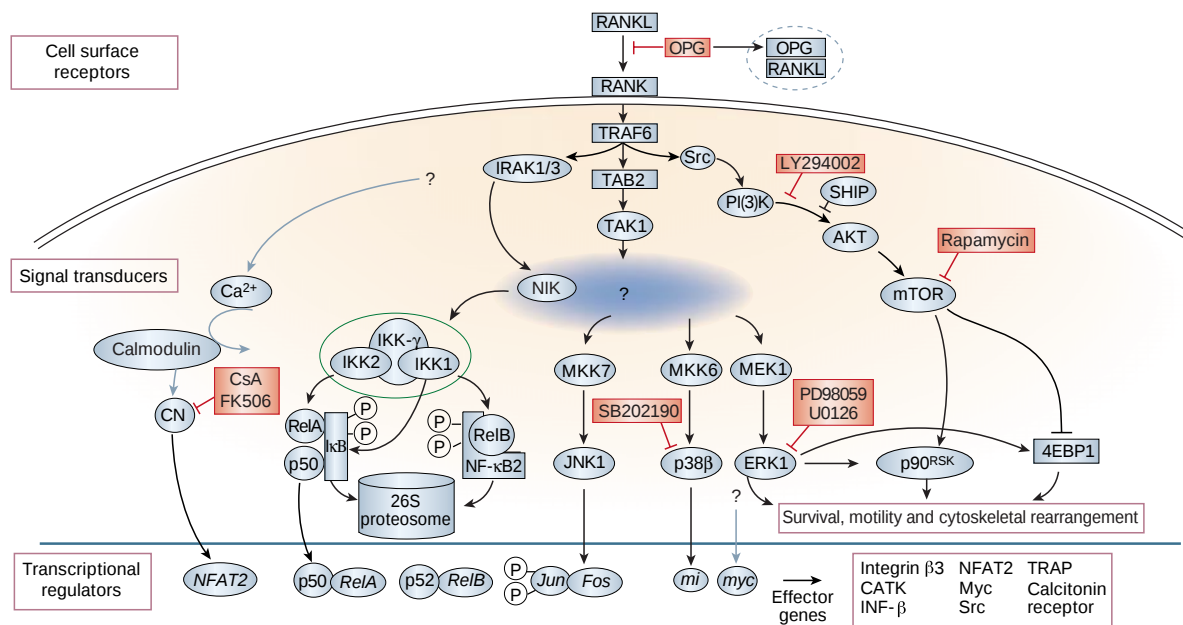


Figure 3 RANK signalling network in osteoclast. Proteins with validated effects involved in RANK signal transduction during osteoclast development and activation, arranged in a signalling cascade from the cytoplasmic membrane to nuclear effectors. RANK and OPG are TNFR receptor-related proteins, and RANKL is a TNF-related

cytokine that interacts specifically with either RANK or OPG. Protein nodes are shown in representations of their functional annotation, and are connected by edges (arrows or lines) that denote functional associations depicted in the literature. Red bars indicate target of known small-molecule inhibitors.

effect is to increase bone resorption. The cytokine IL-4 has also recently been shown to negatively regulate osteoclastogenesis, and this process is controlled downstream by STAT (signal transducer and activator of transcription) signalling within the osteoclast⁶⁰. Finally, the binding of calcitonin to its receptor has long been known to dampen osteoclast activation and is the basis for using it as a therapeutic, although its molecular mechanism of action is unclear.

Hormonal control of bone resorption

Certain hormones, cytokines and humoral factors produced in distant organs can also influence bone density and calcium homeostasis locally by inducing RANKL expression within the bone cells (Fig. 4). Most, if not all, calciotropic hormones and pro-resorptive cytokines have been shown to upregulate messenger RNA expression of RANKL in osteoblast cell lines and primary cell cultures^{32,33}. OPG, which blocks osteoclastogenesis induced by RANKL, can inhibit osteoclast formation and bone resorption induced by treatment with calciotropic factors, indicating that the RANK signalling pathway in osteoclasts integrates diverse humoral signals that regulate bone resorption and calcium homeostasis. Consistent with this hypothesis, RANK-deficient mice are resistant to bone resorption induced by TNF- α , IL-1 β , 1,25(OH) $_2$ vitamin D $_3$ and parathyroid hormone (PTH)-related peptide (PTHrP), which, excluding PTH, are the major calciotropic factors that are known to induce increases in bone resorption and serum hypercalcaemia²⁷. The T cell is also an important source of RANKL in the bone²⁹. Activation of T cells *in vitro* and *in vivo* leads to increased osteoclastogenesis and bone resorption, suggesting that acute and chronic inflammatory states, and certain leukaemias, contribute to pathologic bone loss⁶⁶.

Humoral factors that decrease bone resorption and increase density, such as oestrogens, have a converse effect on the coupling between the osteoblast and osteoclast — OPG expression is increased and/or RANKL expression is decreased, leading to decreased activation of RANK and subsequently the numbers of activated osteoclasts in the bone. Recently the cytokine thrombopoietin, which regulates platelet levels, has also been shown to induce OPG expression in animals,

leading to abnormal increases in bone density⁶⁷. Osteoblast and marrow stromal cells, as well as T cells, therefore act as a key local responding unit within the bone that interpret physiological cues that impact bone remodelling rates, calcium release and bone density (Fig. 4).

Bench to clinic

Except for the recently approved PTH-derived peptide Forteo, current treatments for osteoporosis chiefly retard bone mineral density loss, thereby decreasing the risk for vertebral and long-bone fractures. These compounds target osteoclast-mediated bone resorption and include oestrogens, bisphosphonates and selective oestrogen-receptor modulators (SERMs) (Table 1). Pharmaceutical companies are trying to develop SERMs with improved efficacy and safety profiles, bisphosphonates with more convenient formulations and dosing regimens, and oral forms of calcitonin. Other potential osteoclast targets with chemical inhibitors in preclinical development include the osteoclast-specific protease CATK, the integrin $\alpha_v\beta_3$, and the c-Src tyrosine kinase. All of these osteoclast target molecules are induced during osteoclastogenesis via RANK activation driven by RANKL or RANK-activating antibodies^{7,9}.

The crucial role of the RANKL/RANK/OPG signalling pathways in regulating bone metabolism is underscored by recent findings that genetic mutations that activate RANK or inhibit the RANKL binding properties of OPG in humans are associated with familial forms of hyperphosphatasia and bone abnormalities^{68–72}. Activating mutations in exon 1 of *TNFRSF11A*, the gene encoding RANK, have been found to be associated with osteolytic and non-osteolytic forms of hyperphosphatasia. Expansile skeletal hyperphosphatasia and familial expansile osteolysis are allelic bone diseases caused by overlapping 18-bp and 15-bp tandem duplications in exon 1 of *TNFRSF11A*, respectively, which lengthen the signal peptide of RANK and likely increase its biological activity by sequestering the receptor intracellularly, causing excessive signalling through NF- κ B^{68,69}.

Several mutations in *TNFRSF11B*, the gene encoding OPG, have been found to be associated with idiopathic hyperphosphatasia (also known as juvenile Paget's disease) in humans, an autosomal recessive

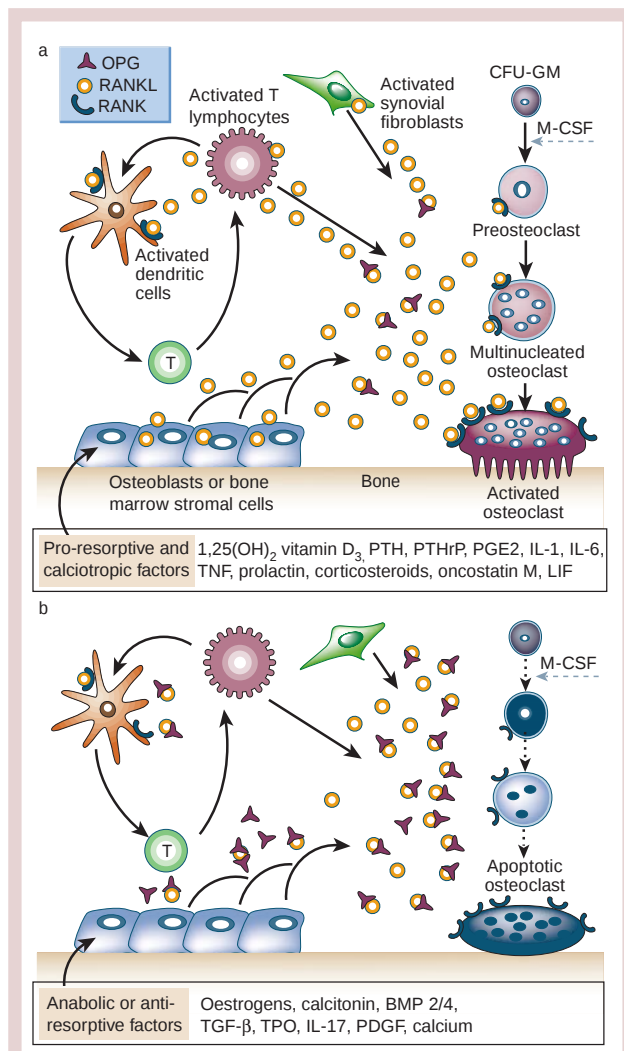


Figure 4 Hormonal control of bone resorption. Schematic representation of the mechanism of action of **a**, pro-resorptive and calcitropic factors; and **b**, anabolic and anti-osteoclastic factors. RANKL expression is induced in osteoblasts, activated T cells, synovial fibroblasts and bone marrow stromal cells, and subsequently binds to its specific membrane-bound receptor RANK, thereby triggering a network of TRAF-mediated kinase cascades that promote osteoclast differentiation, activation and survival. Conversely, OPG expression is induced by factors that block bone catabolism and promote anabolic effects. OPG binds and neutralizes RANKL, leading to a block in osteoclastogenesis and decreased survival of pre-existing osteoclasts.

bone disease characterized by deformities of long bones and kyphosis^{69–71}. Individuals with this disorder exhibit widening of the long-bone diaphyses with a propensity to fracture, accelerated bone turnover, and are typically of short stature. These individuals are phenotypically similar to OPG-knockout mice²⁰.

The observations that mutations in the genes encoding RANK and OPG cause bone diseases of such striking severity in humans suggest that inhibition of RANKL signalling may be a viable therapeutic strategy for treatment of diseases where excessive resorption or remodelling of bone prevail. In animal models, RANKL antagonists that have shown robust activity include full-length OPG as well as the ligand-binding domains of either OPG or RANK fused to Fc region of immunoglobulin-γ²⁴. There is hope that blocking RANKL could preserve bone loss that results from menopause, cancer, inflammation, microgravity/disuse or excesses of either PTH/PTHrP or vitamin D. In the clinic, the Fc-OPG fusion molecules have demonstrated

Table 1 Classes of resorption inhibitors approved for use or in development

Compound class	Type	Mechanism of action
Bisphosphonates	Small molecule	Blocks isoprenylation of Rho and Rap and induces apoptosis in osteoclasts
Amino		Metabolized to cytotoxic ATP analogues
Oestrogens/SERMs	Small molecule	Oestrogen-receptor agonist
RANKL antagonists	Recombinant protein	Blocks RANKL–RANK interactions
α ₂ β ₃ antagonists*	Small molecule	Blocks osteoclast adhesion to bone
Src inhibitors*	Small molecule	Blocks steps leading to osteoclast activation
Cathepsin K inhibitors*	Small molecule	Blocks activity of osteoclast-specific collagenase
Calcitonin*	Peptide	Calcitonin-receptor agonist decreases osteoclast activity

An asterisk indicates those targets that are induced in osteoclasts by RANKL stimulation, and whose expression and activity is decreased by RANKL blockade. Oestrogen replacement therapy downregulates the level of RANKL expression in bone marrow cells in postmenopausal women⁷⁵, and bisphosphonates abrogate the effects of RANK signals that control osteoclast survival and induce cell death⁷⁶.

long-acting suppression of bone-turnover markers with a good safety profile^{73,74}. Recently, a fully human monoclonal antibody directed against RANKL has entered into early development (P. J. Bekker *et al.*, unpublished data). An appealing aspect of the antibody approach is avoidance of either cross-reacting neutralizing OPG or RANK-activating endogenous antibodies that could lead to safety issues. Whether blocking RANKL leads to clinical benefit will be ultimately addressed in clinical trials.

Future directions

Since the discovery of the RANK/RANKL/OPG signalling axis, our understanding of the osteoclast and its role in bone biology has advanced considerably. Based on the literature, a network of over 60 proteins in the osteoclast are associated with pathways regulating the osteoclast, approximately ten times the number identified three years ago. Using expression data and functional assays, it may be feasible to reconstruct the entire signalling network within the osteoclast, to create mechanistic views of diverse osteoclast functions and their impact on bone remodelling and density. Thus, certain areas of osteoclast biology that remain unclear may be elucidated, leading to the development of therapeutics optimized for both safety and efficacy. □

doi:10.1038/nature01658

- Chambers, T. J. Regulation of the differentiation and function of osteoclasts. *J. Pathol.* **192**, 4–13 (2000).
- Teitelbaum, S. L. Bone resorption by osteoclasts. *Science* **289**, 1504–1508 (2000).
- Rodan, G. A. & Martin, T. J. Therapeutic approaches to bone diseases. *Science* **289**, 1508–1514 (2000).
- Khosla, S. Minireview: the OPG/RANKL/RANK system. *Endocrinology* **142**, 5050–5055 (2001).
- Takahashi, N. *et al.* Osteoclast-like cell formation and its regulation by osteotropic hormones in mouse bone marrow cultures. *Endocrinology* **122**, 1373–1382 (1988).
- Yasuda, H. *et al.* Osteoclast differentiation factor is a ligand for osteoprotegerin/osteoclastogenesis-inhibitory factor and is identical to TRANCE/RANKL. *Proc. Natl Acad. Sci. USA* **95**, 3597–3602 (1998).
- Lacey, D. L. *et al.* Osteoprotegerin ligand is a cytokine that regulates osteoclast differentiation and activation. *Cell* **93**, 165–176 (1998).
- Nakagawa, N. *et al.* RANK is the essential signaling receptor for osteoclast differentiation factor in osteoclastogenesis. *Biochem. Biophys. Res. Commun.* **253**, 395–400 (1998).
- Hsu, H. *et al.* Tumor necrosis factor receptor family member RANK mediates osteoclast differentiation and activation induced by osteoprotegerin ligand. *Proc. Natl Acad. Sci. USA* **96**, 3540–3545 (1999).
- Burgess, T. L. *et al.* The ligand for osteoprotegerin (OPGL) directly activates mature osteoclasts. *J. Cell Biol.* **145**, 527–538 (1999).
- Li, Y. P., Chen, W., Liang, Y., Li, E. & Stashenko, P. *Atg6*-deficient mice exhibit severe osteopetrosis due to loss of osteoclast-mediated extracellular acidification. *Nature Genet.* **23**, 447–451 (1999).
- Fuller, K., Murphy, C., Kirstein, B., Fox, S. W. & Chambers, T. J. TNFα potentially activates osteoclasts, through a direct action independent of and strongly synergistic with RANKL. *Endocrinology* **143**, 1108–1118 (2002).
- Rodan, G. A. Bone homeostasis. *Proc. Natl Acad. Sci. USA* **95**, 13361–13362 (1998).
- Jimi, E. *et al.* Osteoclast differentiation factor acts as a multifunctional regulator in murine osteoclast differentiation and function. *J. Immunol.* **163**, 434–442 (1999).
- Lacey, D. L. *et al.* Osteoprotegerin ligand modulates murine osteoclast survival *in vitro* and *in vivo*. *Am. J. Pathol.* **157**, 435–448 (2000).

16. Marks, S. C. Jr Osteoclast biology: lessons from mammalian mutations. *Am. J. Med. Genet.* **34**, 43–53 (1989).
17. McLean, W. & Olsen, B. R. Mouse models of abnormal skeletal development and homeostasis. *Trends Genet.* **17**, S38–S43 (2001).
18. Van Wesenbeeck, L. *et al.* The osteopetrotic mutation *toothless (tl)* is a loss-of-function frameshift mutation in the rat *Csf1* gene: evidence of a crucial role for CSF-1 in osteoclastogenesis and endochondral ossification. *Proc. Natl. Acad. Sci. USA* **99**, 14303–14308 (2002).
19. Takeshita, S. *et al.* SHIP-deficient mice are severely osteopetrotic due to increased numbers of hyper-resorptive osteoclasts. *Nature Med.* **9**, 943–949 (2002).
20. Bucay, N. *et al.* Osteoprotegerin-deficient mice develop early onset osteoporosis and arterial calcification. *Genes Dev.* **12**, 1260–1268 (1998).
21. Wagner, E. F. & Karsenty, G. Genetic control of skeletal development. *Curr. Opin. Genet. Dev.* **5**, 527–532 (2001).
22. Simonet, W. S. *et al.* Osteoprotegerin: a novel secreted protein involved in the regulation of bone density. *Cell* **89**, 309–319 (1997).
23. Yasuda, H. *et al.* Identity of osteoclastogenesis inhibitory factor (OCIF) and osteoprotegerin (OPG): a mechanism by which OPG/OCIF inhibits osteoclastogenesis *in vitro*. *Endocrinology* **139**, 1329–1337 (1998).
24. Morony, S. *et al.* A chimeric form of osteoprotegerin inhibits hypercalcemia and bone resorption induced by IL-1 β , TNF- α , PTH, PTHrP, and 1,25(OH) $_2$ D $_3$. *J. Bone Miner. Res.* **14**, 1478–1485 (1999).
25. Wong, B. R. *et al.* TRANCE is a novel ligand of the tumor necrosis factor receptor family that activates c-jun N-terminal kinase in T cells. *J. Biol. Chem.* **272**, 25190 (1997).
26. Anderson, D. M. *et al.* A homologue of the TNF receptor and its ligand enhance T-cell growth and dendritic-cell function. *Nature* **390**, 175–179 (1997).
27. Li, J. *et al.* RANK is the intrinsic hematopoietic cell surface receptor that controls osteoclastogenesis and regulation of bone mass and calcium metabolism. *Proc. Natl. Acad. Sci. USA* **97**, 1566–1571 (2000).
28. Dougall, W. C. *et al.* RANK is essential for osteoclast and lymph node development. *Genes Dev.* **13**, 2412–2424 (1999).
29. Kong, Y. Y. *et al.* OPGL is a key regulator of osteoclastogenesis, lymphocyte development and lymph-node organogenesis. *Nature* **397**, 315–323 (1999).
30. Hofbauer, L. C. *et al.* The roles of osteoprotegerin and osteoprotegerin ligand in the paracrine regulation of bone resorption. *J. Bone Miner. Res.* **15**, 2–12 (2000).
31. Theill, L. E., Boyle, W. J. & Penninger, J. M. RANK-L and RANK: T cells, bone loss, and mammalian evolution. *Annu. Rev. Immunol.* **20**, 795–823 (2002).
32. Udagawa, N. *et al.* Osteoprotegerin produced by osteoblasts is an important regulator in osteoclast development and function. *Endocrinology* **141**, 3478–3484 (2000).
33. Schoppert, M., Preissner, K. T. & Hofbauer, L. C. RANK ligand and osteoprotegerin: paracrine regulators of bone metabolism and vascular function. *Arterioscler. Thromb. Vasc. Biol.* **22**, 549–553 (2002).
34. Darnay, B. G., Haridas, V., Ni, J., Moore, P. A. & Aggarwal, B. B. Characterization of the intracellular domain of receptor activator of NF- κ B (RANK). Interaction with tumor necrosis factor receptor-associated factors and activation of NF- κ B and c-Jun N-terminal kinase. *J. Biol. Chem.* **273**, 20551–20555 (1998).
35. Galibert, L., Tometsko, M. E., Anderson, D. M., Cosman, D. & Dougall, W. C. The involvement of multiple tumor necrosis factor receptor (TNFR)-associated factors in the signaling mechanisms of receptor activator of NF- κ B, a member of the TNFR superfamily. *J. Biol. Chem.* **273**, 34120–34127 (1998).
36. Lomaga, M. A. *et al.* TRAF6 deficiency results in osteopetrosis and defective interleukin-1, CD40, and LPS signaling. *Genes Dev.* **13**, 1015–1024 (1999).
37. Kobayashi, N. *et al.* Segregation of TRAF6-mediated signaling pathways clarifies its role in osteoclastogenesis. *EMBO J.* **20**, 1271–1280 (2001).
38. Armstrong, A. P. *et al.* A RANK/TRAF6-dependent signal transduction pathway is essential for osteoclast cytoskeletal organization and resorptive function. *J. Biol. Chem.* **277**, 44347–44356 (2002).
39. Ye, H. *et al.* Distinct molecular mechanism for initiating TRAF6 signalling. *Nature* **418**, 443–447 (2002).
40. Franzoso, G. *et al.* Requirement for NF- κ B in osteoclast and B-cell development. *Genes Dev.* **11**, 3482–3496 (1997).
41. Xing, L. *et al.* NF- κ B p50 and p52 expression is not required for RANK-expressing osteoclast progenitor formation but is essential for RANK- and cytokine-mediated osteoclastogenesis. *J. Bone Miner. Res.* **17**, 1200–1210 (2002).
42. Grigoriadis, A. E. *et al.* c-Fos: a key regulator of osteoclast-macrophage lineage determination and bone remodeling. *Science* **266**, 443–448 (1994).
43. Karin, M., Cao, Y., Greten, F. R. & Li, Z.-W. NF- κ B in cancer: from an innocent bystander to major culprit. *Nature Rev. Cancer* **2**, 301–310 (2002).
44. David, J. P., Sabapathy, K., Hoffmann, O., Idarraga, M. H. & Wagner, E. F. JNK1 modulates osteoclastogenesis through both c-Jun phosphorylation-dependent and -independent mechanisms. *J. Cell Sci.* **115**, 4317–4325 (2002).
45. Lee, S. W., Han, S. I., Kim, H. H. & Lee, Z. H. TAK1-dependent activation of AP-1 and c-Jun N-terminal kinase by receptor activator of NF- κ B. *J. Biochem. Mol. Biol.* **35**, 371–376 (2002).
46. Mizukami, J. *et al.* Receptor activator of NF- κ B ligand (RANKL) activates TAK1 mitogen-activated protein kinase kinase kinase through a signaling complex containing RANK, TAB2, and TRAF6. *Mol. Cell. Biol.* **22**, 992–1000 (2002).
47. Oda, H., Nakamura, K. & Tanaka, S. Possible involvement of I κ B kinase 2 and MKK7 in osteoclastogenesis induced by receptor activator of nuclear factor κ B ligand. *J. Bone Miner. Res.* **17**, 612–621 (2002).
48. Li, X. *et al.* p38 MAPK-mediated signals are required for inducing osteoclast differentiation but not for osteoclast function. *Endocrinology* **143**, 3105–3113 (2002).
49. Matsumoto, M., Sudo, T., Saito, T., Osada, H. & Tsujimoto, M. Involvement of p38 mitogen-activated protein kinase signaling pathway in osteoclastogenesis mediated by receptor activator of NF- κ B ligand (RANKL). *J. Biol. Chem.* **275**, 31155–31161 (2000).
50. Mansky, K. C., Sankar, U., Han, J. & Ostrowski, M. C. Microphthalmia transcription factor is a target of the p38 MAPK pathway in response to receptor activator of NF- κ B ligand signaling. *J. Biol. Chem.* **277**, 11077–11083 (2002).
51. Wei, S., Wang, M. W., Teitelbaum, S. L. & Ross, F. P. Interleukin-4 reversibly inhibits osteoclastogenesis via inhibition of NF- κ B and mitogen-activated protein kinase signaling. *J. Biol. Chem.* **277**, 6622–6630 (2002).
52. Hotokezaka, H. *et al.* U0126 and PD98059, specific inhibitors of MEK, accelerate differentiation of RAW264.7 cells into osteoclast-like cells. *J. Biol. Chem.* **277**, 47366–47372 (2002).
53. Wong, B. R. *et al.* TRANCE, a TNF family member, activates Akt/PKB through a signaling complex involving TRAF6 and c-Src. *Mol. Cell* **4**, 1041–1049 (1999).
54. Shui, C., Riggs, B. L. & Khosla, S. The immunosuppressant rapamycin, alone or with transforming growth factor- β , enhances osteoclast differentiation of RAW264.7 monocyte-macrophage cells in the presence of RANKL-ligand. *Calcif. Tissue Int.* **71**, 437–446 (2002).
55. Cappellen, D. *et al.* Transcriptional program of mouse osteoclast differentiation governed by the macrophage colony-stimulating factor and the ligand for the receptor activator of NF κ B. *J. Biol. Chem.* **277**, 21971–21982 (2002).
56. Ishida, N. *et al.* Large scale gene expression analysis of osteoclastogenesis *in vitro* and elucidation of NFAT2 as a key regulator. *J. Biol. Chem.* **277**, 41147–41156 (2002).
57. Takayanagi, H. *et al.* Induction and activation of the transcription factor NFATc1 (NFAT2) integrate RANKL signaling in terminal differentiation of osteoclasts. *Dev. Cell* **3**, 889–901 (2002).
58. Battaglini, R. *et al.* c-myc is required for osteoclast differentiation. *J. Bone Miner. Res.* **17**, 763–773 (2002).
59. Mak, T. W. & Yeh, W. C. A block at the toll gate. *Nature* **418**, 835–836 (2002).
60. Arai, F. *et al.* Commitment and differentiation of osteoclast precursor cells by the sequential expression of c-Fms and receptor activator of nuclear factor κ B (RANK) receptors. *J. Exp. Med.* **190**, 1741–1754 (1999).
61. Yan, T., Riggs, B. L., Boyle, W. J. & Khosla, S. Regulation of osteoclastogenesis and RANK expression by TGF- β 1. *J. Cell. Biochem.* **83**, 320–325 (2001).
62. Ma, Y. L. *et al.* Catabolic effects of continuous human PTH (1–38) *in vivo* is associated with sustained stimulation of RANKL and inhibition of osteoprotegerin and gene-associated bone formation. *Endocrinology* **142**, 4047–4054 (2001).
63. Takayanagi, H. *et al.* RANKL maintains bone homeostasis through c-Fos-dependent induction of *interferon- β* . *Nature* **416**, 744–749 (2002).
64. Hayashi, T., Kaneda, T., Toyama, Y., Kumegawa, M. & Hakeda, Y. Regulation of receptor activator of NF- κ B ligand-induced osteoclastogenesis by endogenous interferon- β (INF- β) and suppressors of cytokine signaling (SOCS). The possible counteracting role of SOCS in IFN- β -inhibited osteoclast formation. *J. Biol. Chem.* **277**, 27880–27886 (2002).
65. Takayanagi, H. *et al.* T-cell-mediated regulation of osteoclastogenesis by signalling cross-talk between RANKL and IFN- γ . *Nature* **408**, 600–605 (2000).
66. Kong, Y. Y. *et al.* Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand. *Nature* **402**, 304–309 (1999).
67. Chagraoui, H. *et al.* Stimulation of osteoprotegerin production is responsible for osteosclerosis in mice overexpressing TPO. *Blood* (in the press).
68. Hughes, A. E. *et al.* Mutations in *TNFRSF11A*, affecting the signal peptide of RANK, cause familial expansile osteolysis. *Nature Genet.* **24**, 45–48 (2000).
69. Whyte, M. P. & Hughes, A. E. Expansile skeletal hyperphosphatasia is caused by a 15-base pair tandem duplication in *TNFRSF11A* encoding RANK and is allelic to familial expansile osteolysis. *J. Bone Miner. Res.* **17**, 26–29 (2002).
70. Whyte, M. P. *et al.* Osteoprotegerin deficiency and juvenile Paget's disease. *N. Engl. J. Med.* **347**, 175–184 (2002).
71. Cundy, T. *et al.* A mutation in the gene *TNFRSF11B* encoding osteoprotegerin causes an idiopathic hyperphosphatasia phenotype. *Hum. Mol. Genet.* **11**, 2119–2127 (2002).
72. Langdahl, B. L., Carstens, M., Stenkaer, L. & Eriksen, E. F. Polymorphisms in the osteoprotegerin gene are associated with osteoporotic fractures. *J. Bone Miner. Res.* **17**, 1245–1255 (2002).
73. Bekker, P. J. *et al.* The effect of a single dose of osteoprotegerin in postmenopausal women. *J. Bone Miner. Res.* **16**, 348–360 (2001).
74. Body, J. J. *et al.* A Phase I study of AMG-0007, a recombinant osteoprotegerin construct, in patients with multiple myeloma or breast carcinoma related bone metastases. *Cancer* **97**(Suppl.), 887–892 (2003).
75. Eghbali-Fatourehchi, G. *et al.* Role of RANK ligand in mediating the increased bone resorption in early postmenopausal women. *J. Clin. Invest.* **111**, 1221–1230 (2003).
76. Kim, Y. H., Kim, G. S. & Jeong-Hwa, B. Inhibitory action of bisphosphonates on bone resorption does not involve the regulation of RANKL and OPG expression. *Exp. Mol. Med.* **34**, 145–151 (2002).
77. Mizuno, A. *et al.* Transgenic mice overexpressing soluble osteoclast differentiation factor (sODF) exhibit severe osteoporosis. *J. Bone Miner. Metab.* **20**, 337–344 (2002).

The genetic basis for skeletal diseases

Elazar Zelzer & Bjorn R. Olsen

Harvard Medical School, Department of Cell Biology, 240 Longwood Avenue, Boston, Massachusetts 02115, USA
(e-mail: bjorn_olsen@hms.harvard.edu)

We walk, run, work and play, paying little attention to our bones, their joints and their muscle connections, because the system works. Evolution has refined robust genetic mechanisms for skeletal development and growth that are able to direct the formation of a complex, yet wonderfully adaptable organ system. How is it done? Recent studies of rare genetic diseases have identified many of the critical transcription factors and signalling pathways specifying the normal development of bones, confirming the wisdom of William Harvey when he said: "nature is nowhere accustomed more openly to display her secret mysteries than in cases where she shows traces of her workings apart from the beaten path".

Genetic studies of diseases that affect skeletal development and growth are providing invaluable insights into the roles not only of individual genes, but also of entire developmental pathways. Different mutations in the same gene may result in a range of abnormalities, and disease 'families' are frequently caused by mutations in components of the same pathway. Correlating these clinical phenotypes with the identified molecular alterations provides a sensitive method for analysing structure–function relationships.

In many cases the alterations are in molecules that — based on biochemical analyses — could not have been reasonably suspected to be important in bone biology. Such surprising discoveries are opening doors to new research areas. Because the gene mutations are often not simple loss-of-function mutations, they provide the impetus for introducing specific mutations into mice to generate models for studies of pathogenesis and fundamental aspects of cell differentiation and function.

Given the space limitations of this review, we have focused on examples that illustrate how studies of disease families are contributing to the understanding of regulatory pathways. We have excluded diseases (such as disorders of extracellular matrix components and various cranio-synostosis syndromes) that have been discussed in recent reviews^{1–3}. Human studies provide important information that complements data from experimental analyses of gene function in animals and cells. Although experimental studies are clearly of enormous importance, examining the phenotypic consequences of gene mutations in inbred mouse strains provides a view of gene function under conditions of minimal genetic and environmental variability. In contrast, studies of gene mutations in human families and populations can provide insights into adaptability of genetic mechanisms and biological systems within a complex environment. A full understanding of the genetic mechanisms for bone assembly, growth and function provided by evolution will be possible only by utilizing information from all these types of studies. This requires open communication between clinical geneticists, biochemists, and cell and developmental biologists, and we hope this brief review will stimulate that communication.

Classification of skeletal diseases

The vertebrate skeleton is formed by mesenchymal cells condensing into tissue elements outlining the pattern of future bones (the patterning phase). This is followed by

differentiation to cartilage cells (chondrocytes) or bone cells (osteoblasts) within the condensations. Subsequent growth during the organogenesis phase generates cartilage models (anlagen) of future bones (as in limb bones) or membranous bones (as in the cranial vault) (Fig. 1). The cartilage anlagen are replaced by bone and marrow in a process called endochondral ossification. Finally, a process of growth and remodelling after birth (in a growth and maintenance phase) results in a skeleton that is well adapted to its function as an organ not only for movement, but also for support and protection of internal organs, blood cell production and regulation of calcium homeostasis.

Mutations in early patterning genes cause disorders called dysostoses; these affect only specific skeletal elements, leaving the rest of the skeleton largely unaffected. In contrast, mutations in genes that are involved primarily in organogenesis cause disorders called osteochondrodysplasias, which affect the development and growth of most skeletal elements in a generalized fashion⁴. Many genes have important functions in both these processes such that some inherited disorders can display features of both dysostosis and osteochondrodysplasia⁵. Genes used during skeletal development may also be important in other organogenetic events so that when mutated, the resulting skeletal genetic defects are part of a syndrome that may also include defects in nonskeletal tissues.

Disorders of bone and cartilage cell differentiation

Disorders affecting differentiation of cartilage and bone cells have features of both dysostosis and osteochondrodysplasia. Studies of two such disorders, campomelic dysplasia (CD; Mendelian Inheritance in Man (MIM) database no. 114290) and cleidocranial dysplasia (CCD; MIM 119600), have led to insights into the transcriptional machinery responsible for differentiation of chondrocytes and osteoblasts (Fig. 2).

The study of the dominant disorder CD represents an excellent example of how identification of human disease genes can lead to fundamental insights into the cellular basis for skeletal development. The discovery that loss-of-function mutations in the transcription factor SOX9 cause CD was followed by studies that comprise most of our current knowledge of how chondrocytic differentiation is transcriptionally regulated^{6,7}. Babies with CD usually die during the perinatal period or in early infancy with skeletal abnormalities that strongly suggest a defect in cartilage, as well as in patterning of specific bones. Airway cartilage is defective, the rib cage is small with a decreased number of ribs, the vertebral column is abnormal, and the base of the skull (formed

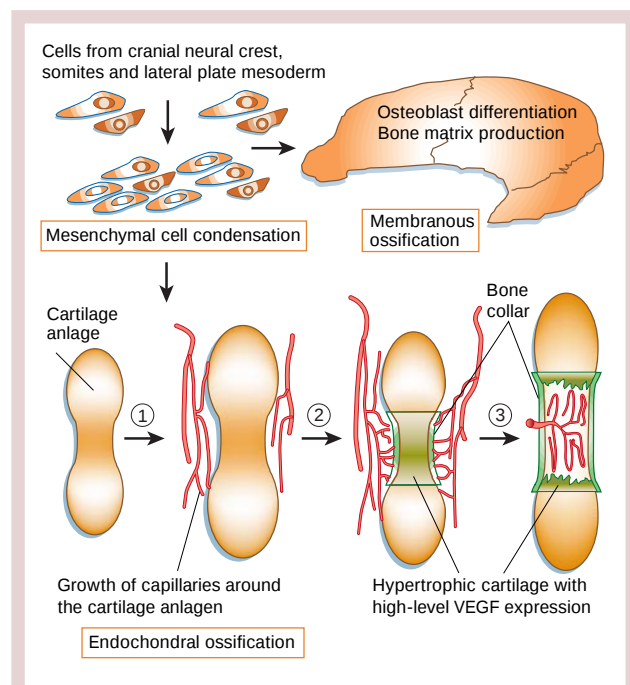


Figure 1 Skeletal development. Cells from cranial neural crest, somites and lateral plate mesoderm form mesenchymal condensations at sites of future bones. In membranous ossification, differentiation of mesenchymal cells to osteoblasts and production of bone matrix occurs directly, whereas in endochondral ossification, differentiation to chondrocytes and formation of cartilage models (anlagen) occurs first, followed by replacement of the models by bone. In step one, capillaries grow around cartilage anlagen. In step two, cells around the anlage differentiate to osteoblasts and produce a collar of bone around the middle of the cartilage. Chondrocytes in the centre of the cartilage mature to hypertrophy and express high levels of vascular endothelial growth factor (VEGF), a factor needed for invasion of capillaries into the cartilage. In step three, hypertrophic cartilage is replaced by marrow and bone and growth plates are formed (see review in this issue by Kronenberg, page 332).

through endochondral ossification) is small. Many of the generalized defects are similar to defects caused by mutations in the cartilage collagens II and XI (ref. 1).

Heterozygous *Sox9*-null mice (*Sox9*^{+/-}) display most of the abnormalities seen in CD patients⁸. Mutant mice die at birth with cleft palate and abnormal bending of endochondral bones in the limbs. Changes in the growth plates of *Sox9*^{+/-} embryos suggest that *Sox9* may inhibit the maturation of proliferating cells to hypertrophic chondrocytes. Inactivation of *Sox9* in limb buds before mesenchymal condensations are formed results in complete absence of cartilage and bone⁹. In contrast, inactivation after condensations are formed results in defects similar to those reported for mice with inactivated *Sox5* and *Sox6* alleles¹⁰. Coupled with studies showing that *Sox9* can bind to and activate the promoters/enhancers of chondrocyte-specific genes¹¹, these studies show that *Sox9* is an important regulator of differentiation and maturation of chondrocytes (Fig. 2).

Cleidocranial dysplasia is a dominant condition characterized by defects in both endochondral and intramembranous bone formation, such as small or absent clavicles, fontanelles that do not fill in with bone, supernumerary teeth and short stature. The discovery of the transcription factor CBFA1/RUNX2 (core binding factor a1/runt homeodomain protein 2) as a critical regulator of bone formation and the cause of CCD in both humans and mice represents one of those key events when independent efforts in several laboratories lead to a confluence of discoveries highlighting one specific molecular process. In this case, several parallel avenues of research, namely,

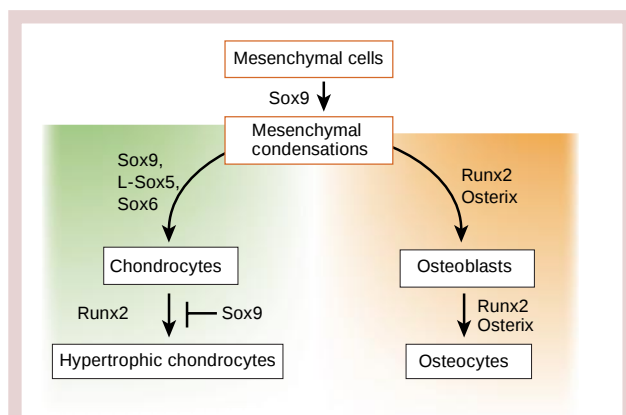


Figure 2 Differentiation factors in chondrocytic and osteoblastic differentiation. *Sox9*, together with *Sox5* and *Sox6*, regulates differentiation of chondrocytes. *Runx2* and *Osterix* control osteoblastic differentiation. Maturation of chondrocytes to hypertrophy is controlled positively by *Runx2* and negatively by *Sox9*. As *Runx2* has a role in both chondrocytic and osteoblastic differentiation, *Osterix* is likely the factor that provides specificity in osteoblastic differentiation.

on transcriptional regulation of the bone protein osteocalcin¹², on mice with inactivated *Runx2* alleles aimed at determining a possible role for this factor in T-cell development^{13,14}, and on mutational analyses of mice and human patients with CCD¹⁵, all contributed to the conclusion that RUNX2 is a critical transcription factor for osteoblastic differentiation and chondrocytic maturation (Fig. 2). Patients with CCD are heterozygous for loss-of-function mutations in RUNX2.

In the complete absence of *Runx2* activity, as in *Runx2*-deficient mice, the entire skeleton consists only of cartilage, most of which is composed of resting or proliferating chondrocytes, or connective tissue membranes with no ossification¹⁶. In some limb cartilages of the endochondral skeleton, hypertrophy of chondrocytes is absent. But *Runx2* cannot be absolutely necessary for chondrocytic maturation, as cartilage anlagen do show hypertrophy in the distal limbs (tibia, fibula, radius and ulna)¹⁶. The reason for this difference in sensitivity of chondrocytic hypertrophy to loss of *Runx2* function in different bones is unknown. As two other members of the Runx family, *Runx1* and *Runx3*, are expressed in growth-plate chondrocytes, it is possible that they can compensate for the loss of *Runx2* in distal limbs. All three Runx factors form heterodimers with a common subunit, core-binding factor β (Cbf β), recently demonstrated to be crucial for bone formation^{17–19}. In the absence of *Runx2*, heterodimers of Cbf β with *Runx1* and/or *Runx3* may be sufficient to allow cartilage hypertrophy in distal limbs.

The detailed pathways downstream of *Runx2* that control osteoblastic differentiation are not known, but one critical mediator of *Runx2* action is *Osterix* (Osx), a zinc finger-transcription factor expressed in osteoblasts²⁰. *Osx*-deficient mice exhibit a complete block in osteoblastic differentiation but no overt abnormality in chondrocyte differentiation and maturation (Fig. 2).

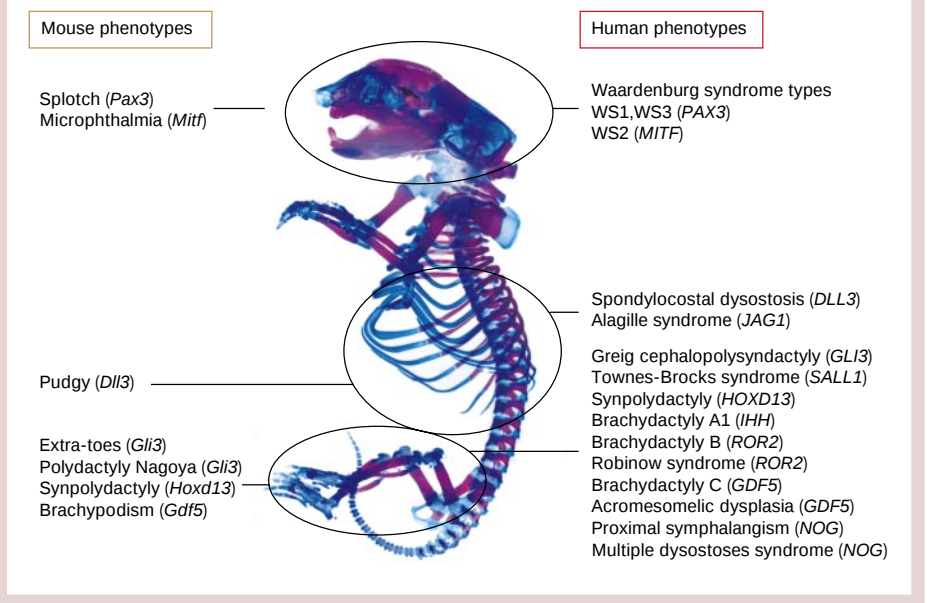
Dysostotic defects of patterning and cell differentiation

Dysostoses can be divided into three groups, reflecting the different origins of the progenitor cells and differences in the developmental processes by which skeletal elements are formed in the craniofacial, axial and limb skeleton (Fig. 3).

Craniofacial patterning and differentiation defects

Craniofacial dysostoses include abnormalities in migration and/or differentiation of cranial neural crest cells (see review in this issue by Helms and Schneider, page 326). Analyses of variants of one of these disorders, Waardenburg syndrome, together with relevant mouse

Figure 3 Mouse and human phenotypes caused by mutations affecting skeletal patterning and differentiation. The grouping of the disorders reflects the different origins of the progenitor cells in the craniofacial (cranial neural crest), axial (somites) and limb skeleton (lateral plate mesoderm). Only disorders discussed in the text are listed. The responsible genes are in parentheses after the names of the syndromes.



models, have greatly advanced the understanding of regulation of neural crest cell differentiation, migration and interactions and illustrate how disease families can help elucidate important developmental pathways. The skeletal manifestations of Waardenburg syndrome include patterning defects in craniofacial and limb bones. In addition, one of the responsible genes encodes a component of a transcription factor network involved in differentiation of multinucleated bone-resorbing cells, osteoclasts, from mononuclear progenitors. For this reason we briefly describe three of the four clinical types (WS1–WS4) of Waardenburg syndrome, which is characterized by deafness and pigmentation abnormalities in variable combination with skeletal anomalies and congenital megacolon.

WS1 (MIM 193500) and WS3 (MIM 148820) are caused by mutations in *PAX3*, a member of a family of transcription factors containing two interdependent DNA-binding domains — a homeodomain and a paired domain²¹. *PAX3* is expressed in somites and the dorsal region of the neural tube where neural crest cells start their migration²². In humans, several different *PAX3* mutations can cause abnormalities in craniofacial bone and soft tissues, spina bifida and cleft lip/palate. *PAX3* is also an important regulator of the migration of precursor cells for muscle into the developing limbs²³. Patients with WS3 show limb defects such as fusion of wrist bones, syndactyly, and small or missing phalangeal bones. In the mouse, point mutations or deletions in *Pax3* cause the Splotch (Sp) phenotype²⁴.

One of the targets of *PAX3* is *Microphthalmia* (*MITF*), a member of the *MITF*-TFE family of basic helix-loop-helix- and leucine zipper-containing transcription factors²⁵. Mutant *PAX3* proteins associated with WS1 fail to activate the *MITF* promoter. Heterozygous mutations in *MITF* are found in many patients with WS2 (MIM 193510)²⁶. Mice with mutations in *Mitf* (*mi*) serve as models for WS2 (ref. 27). In addition, and unlike the situation in humans, some of the strong dominant mutations in *Mitf* cause increased deposition of bone (osteopetrosis) in homozygous animals. This skeletal abnormality is due to a defect in the formation of bone-resorbing osteoclasts, caused by a block in colony stimulating factor-1 (CSF-1)-mediated signalling. Stimulation of mononuclear osteoclast progenitor cells by CSF-1 (in combination with RANK ligand, RANKL) leads to phosphorylation of *Mitf* and its homologue *Tfe3*. The two modified factors in turn associate with a transcriptional co-activator (p300/CBP) to stimulate formation of multinucleated osteoclasts²⁸ (Fig. 4). Osteoclastogenesis is normal in mice lacking either *Mitf* or *Tfe3*, but the combined loss of both transcription factors, results in

severe lack of osteoclasts²⁹. In addition to reduced osteoclast numbers in mice with dominant mutations in *Mitf*, expression of cathepsin K, essential for the ability of osteoclasts to degrade the organic matrix in bone, is also reduced. This is a consequence of the inability of mutant *Mitf* to activate the cathepsin K promoter³⁰.

Axial patterning and differentiation defects

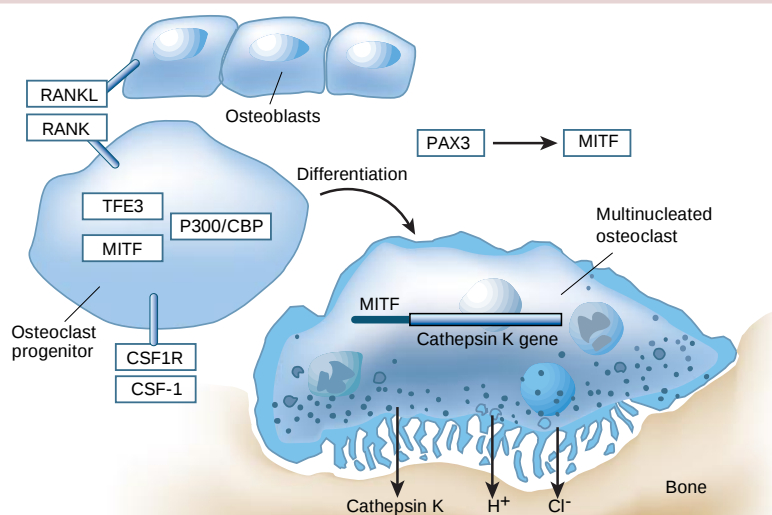
Mutations associated with dysostoses of the axial skeleton affect somite development and cause abnormalities in the segmentation of the vertebral column (spondylocostal dysostosis; MIM 277300) and formation of caudal structures (sacral agenesis). Numerous studies show the importance of Notch signalling for segmentation of mesoderm into somites. A 'segmentation clock', involving the transcription factor *Hairy 1*, drives the expression of the glycosyl transferase *Lunatic fringe* (*Lfng*) and other components of the Notch–Delta pathway^{31,32}. Mutations in the Notch ligand *Delta-like3* (*Dll3*) cause disruption of the segmentation clock both in humans and mice. In families with a recessive form of spondylocostal dysostosis, associated with loss-of-function mutations in *DLL3*, affected individuals have short stature and vertebral and rib abnormalities³³. Severe vertebral and rib deformities are also seen in an X-ray-induced mouse mutant, *pudgy*, caused by disruption of *Dll3* function³⁴, and mice that are homozygous for targeted *Dll3*-null alleles have vertebral defects caused by perturbation of somite formation³⁵. Mice with inactivated *Lfng* alleles display a phenotype similar to *pudgy*^{36,37}.

Mutations resulting in loss of function of the Notch ligand *Jagged 1* (*JAG1*) are associated with the dominant Alagille syndrome (MIM 118450) in humans^{38,39}. Affected individuals have abnormalities of the liver, heart, eye, facial structures, and other parts of the skeleton. Short distal phalangeal bones may be present, but the most striking anomaly is abnormally shaped vertebral bodies (butterfly vertebrae). In mice, *Jagged* does not seem to be required for somite formation. *Jag1*^{+/-} mice have ocular defects similar to eye anomalies sometimes seen in Alagille syndrome patients. Mice that are doubly heterozygous for a *Jag1*-null allele and a hypomorphic *Notch2* allele exhibit all the nonskeletal components of Alagille syndrome, but the bone defects are not present^{40,41}.

Limb patterning and cell differentiation defects

Dysostoses of the limb skeleton include abnormalities of patterning of hands and feet and disorders of growth and joint formation. Many of the syndromes (polydactylies and syndactylies) are caused by

Figure 4 Genes causing different types of Waardenburg syndrome and their role in osteoclast differentiation. As described in the text, mutations in the genes encoding PAX3 and MITF cause WS1–WS3. MITF is a critical transcription factor for differentiation of osteoclast progenitors to multinucleated osteoclasts and for activation of the gene encoding cathepsin K, a secreted enzyme required for efficient degradation of bone matrix. As MITF is a downstream target of PAX3 in neural crest-derived cells, it is possible that PAX3 may also have a role in osteoclastogenesis.



mutations in components of the Sonic hedgehog pathway, reflecting the critical importance of Sonic hedgehog for patterning of the limb skeleton. Secreted by cells in the posterior zone of polarizing activity in the growing limb bud, Sonic hedgehog controls the expression of transcription factors such as *Gli3* and *Sall1*, and the nested expression of members of the *Hoxd* cluster (see review in this issue by Mariani and Martin, page 319). In Greig cephalopolysyndactyly (MIM 175700), deletions or truncations in the transcriptional repressor *GLI3* cause broad thumbs, polydactyly and syndactyly, and craniofacial anomalies⁴². In Townes-Brocks syndrome (MIM 107480), mutations in the gene encoding *SALL1*, a zinc-finger transcription factor, result in anomalies (craniofacial, hand, renal and anal) in regions where hedgehog signalling is important during development⁴³. A striking illustration of the importance of *Hoxd* gene expression is provided by synpolydactyly (MIM 186000). In individuals with in-frame expansions of a polyaniline tract in the amino-terminal region of *HOXD13*, there is syndactyly between fingers three and four and duplication of a finger in the syndactylous web in the hand⁴⁴. Good mouse models for these patterning disorders include extra-toes (*xt*)⁴⁵ and polydactyly Nagoya (*Pdn*)⁴⁶, both caused by *Gli3* mutations, and synpolydactyly (*spdh*)⁴⁷, caused by mutations in *Hoxd13*.

Patterning genes control the formation and growth of individual skeletal elements in the limbs. For example, analyses of disorders that cause brachydactylies (shortening of fingers and toes) have identified Indian hedgehog (*Ihh*), receptor tyrosine kinase *ROR2*, and members of the bone morphogenetic protein (BMP) family as important downstream targets. Brachydactylies are divided into several types depending on which long bones and/or digits are affected.

Type A1 brachydactyly (MIM 112500), characterized by short middle phalanges of all digits, short thumbs and short stature, is of particular interest in that it was the first disorder to be described (by S. W. Farabee in his Harvard University Ph.D. thesis in 1903) as an autosomal dominant Mendelian trait. In several families, mutations have been identified in *Ihh*, at residues thought to be important for binding of *Ihh* to its receptor *Patched*⁴⁸. Many of the features of the phenotype can be explained based on the role of *Ihh* in stimulating cell proliferation in growth-plate cartilage.

Brachydactyly type B (MIM 113000) patients, who have normal thumbs, but short fingers II–V, have dominant loss-of-function mutations in *ROR2* (ref. 49). That *ROR2* has an important role in

endochondral bone formation is evident from studies of *Ror2* expression during mouse development and the phenotype of mice that are homozygous for *Ror2*-null alleles⁵⁰. *Ror2* is expressed in chondrocytes and perichondrial cells, and homozygous *Ror2*-null mice die around birth, with shortened snouts, limbs and tails, and cleft palate. Intriguingly, the severity of the cartilage defects increases in the distal parts of the limbs. Middle phalangeal bones are missing; a feature that is similar to what is seen in brachydactyly type B. Homozygosity for loss-of-function mutations in *ROR2* has been demonstrated in several families with recessive Robinow syndrome (MIM 180700). Affected individuals exhibit severe short-limbed dwarfism, rib, vertebral and other abnormalities and brachydactyly^{51,52}.

Dominantly inherited brachydactyly type C (MIM 113100), characterized by shortening of fingers II, III and V, short metacarpals and occasional fusion of the proximal finger joints, is caused by loss-of-function mutations in *GDF5/CDMP-1* (growth and differentiation factor 5/cartilage-derived morphogenetic protein-1), a member of the BMP/transforming growth factor- β family of cytokines⁵³. Homozygosity for a null mutation in *Gdf5* results in mice with short legs (brachypodism)⁵⁴. In humans, homozygosity for loss-of-function mutations in *GDF5* causes Hunter-Thompson type acromesomelic dysplasia (MIM 201250), a disorder with severe shortening of arms and legs and small digits, but with normal craniofacial and axial skeletons⁵⁵. As is the case with *ROR2* mutations, there is a proximal–distal gradient of increasing severity in the limbs.

Proximal symphalangism (MIM 185500), occasionally seen as part of brachydactyly type C, can also be inherited as a distinct dominant disorder, often associated with early onset deafness. Identification of mutations in *NOG* — the secreted BMP-antagonist *Noggin* — explains the similarity to brachydactyly type C⁵⁶. *Noggin* mutations have also been found in multiple dysostoses syndrome (MIM 186500), a disorder characterized by fusion of joints in hands and feet as well as elbows, hips and intervertebral joints⁵⁶. Fusion of limb bones is among the many defects in mice homozygous for *Noggin*-null alleles⁵⁷.

Osteochondrodysplastic growth-plate defects

Osteochondrodysplasias represent a diverse group of malformations. Because they are caused by mutations in genes that affect the function of cartilage and bone tissues, it is not surprising that a large number of these disorders result in abnormal growth plate cartilage with consequent dwarfism.

In growth plates, proliferation and maturation of chondrocytes and replacement of hypertrophic cartilage by bone marrow and bone is a highly regulated process. As discussed elsewhere in this issue by Kronenberg (page 332), Indian hedgehog (Ihh) and parathyroid hormone-like hormone (Pthlh, also called parathyroid hormone-related peptide or PTHrP) form a negative feedback loop that controls differentiation of chondrocytes to hypertrophy. In humans, a dominant disorder, Jansen's metaphyseal chondrodysplasia (MIM 156400), characterized by short limbs and increased bone turnover, is caused by mutations resulting in activation of the receptor (PTHR) for PTHLH⁵⁸. Mutations that impair PTHR function result in Blomstrand's chondrodysplasia (MIM 215045), a recessive disorder characterized by early death and advanced bone maturation⁵⁹. Mutations in *PTHR* have also been identified in individuals with benign cartilage tumours called enchondromas⁶⁰. These can occur as solitary lesions or as multiple lesions in enchondromatosis (MIM 166000). The tumours are usually found in close proximity to growth plates. Consequently, it has been suspected that they result from abnormal regulation of proliferation and differentiation of chondrocytes. Hopyan *et al.*⁶⁰ have recently found a mutation in the extracellular domain of PTHR in two patients with enchondromatosis.

Ihh is a positive regulator of cell proliferation within growth plates (see review by Kronenberg, page 332). In contrast, fibroblast growth factor receptor 3 (Fgfr3) provides negative control. The first demonstration that Fgfr3 signalling is crucial in growth-plate physiology came from reports that mutations in *FGFR3* were responsible for several forms of dwarfism in humans^{61–64}. All the mutations, including those causing the most common form of the disorder, achondroplasia (MIM 100800), result in various levels of activation of FGFR3 signalling⁶⁵. Several mouse models have been made and these show phenotypes that correlate well with what is seen in the human syndromes. Nevertheless, much remains to be learned, as the genotype/phenotype correlation in disorders caused by *FGFR3* mutations is by no means clear. Not only do mutations located in different domains of the receptor cause different phenotypes, but different mutations in the same codon can also cause clinically distinct disorders. Thus, mutations in what is normally a lysine codon within the cytoplasmic tyrosine kinase activation loop of FGFR3 can result in three different syndromes — thanatophoric dysplasia (MIM 187600) when the lysine residue is replaced by glutamic acid or methionine, SADDAN syndrome (MIM 134934) when the lysine is replaced by methionine, and hypochondroplasia (MIM 146000) when the same lysine residue is replaced by asparagine or glutamine.

Within growth plates, expression of Fgfr3 is restricted to proliferating and hypertrophic zones. Because mice with inactivated *Fgf18* alleles have growth-plate alterations that are similar to those of mice with inactivated *Fgfr3* alleles^{66–68}, it has been suggested that Fgf18 is the physiological ligand for Fgfr3. How signalling through Fgfr3 by Fgf18 (or other FGFs) interacts with regulation of growth-plate activities by Ihh/Pthlh is not fully understood. However, recent experiments suggest that FGF signalling acts upstream of Ihh to regulate the onset of chondrocyte differentiation to hypertrophy⁶⁹.

Further insights into growth-plate regulation are provided by studies of Leri-Weill dyschondrosteosis (MIM 127300). Loss-of-function mutations in one allele of the transcription factor *SHOX* (for short stature homeobox gene), located at the tip of both X and Y (sex) chromosomes⁷⁰, results in a phenotype of short stature with shortening of the forearms and lower legs. Other loss-of-function mutations in *SHOX* have also been demonstrated in individuals with so-called idiopathic short stature and *SHOX* is probably responsible for the short stature in patients with Turner syndrome (absence of an X chromosome in females)⁷¹. Langer mesomelic dysplasia (MIM 249700), characterized by severe short stature with marked shortening of the limbs, is caused by loss-of-function mutations in *SHOX* on both sex chromosomes⁷².

Histological study of abnormal growth plates that were surgically removed from the forearm of two patients with dyschondrosteosis

showed shorter than normal columns of chondrocytes in the proliferating zone⁷³. Within the columns, the chondrocytes were stacked side-by-side (not on top of each other as in normal growth plates) and formed nests of cells at various stages of maturation. *SHOX* activity may therefore control the stacking of chondrocytes within the proliferative zones of growth plates and maturation to hypertrophy. Also of interest is the outcome of an unusual 'experiment of nature'. In two families where two siblings with Langer mesomelic dysplasia each married a spouse with achondroplasia or hypochondroplasia, two children had achondroplasia or hypochondroplasia combined with Leri-Weill dyschondrosteosis, whereas two other children had Leri-Weill dyschondrosteosis⁷⁴. A comparison of the affected individuals showed that the different phenotypes, when combined in the same individual, seemed to be milder than what one would expect if they would be additive. It is therefore possible that *SHOX* and *FGFR3* are components of partially overlapping pathways within growth plates. For example, *SHOX* could act downstream of Pthlh/Ihh to control chondrocyte proliferation and maturation.

Future prospects

This brief review has described a small subgroup of a vast number of genetic diseases affecting the skeletal system. Other diseases could have been selected, but the main point to be made would not have been much different. These genetic disorders, apart from their compelling stories of human suffering and triumph, represent 'experiments of nature' with great potential for advancing the understanding of mechanisms underlying skeletal development.

With this in mind, where do we go from here? What are the significant questions in skeletal biology and what can studies of human genetic diseases contribute to the search for the answers? Several important questions come to mind, such as what are the mechanisms that specify the architecture of each and every bone in the skeleton, the processes of joint formation, and the proper insertion of ligaments and tendons? What are the cellular interactions that supply bones with blood vessels and nerves, and how do vessels and nerves support the growth and function of bones? A particularly important question is how cells build and remodel bones in postnatal life. Scientists have for centuries been impressed with the ability of bone cells to produce bone along trajectories of mechanical stress or in locations where maximal strength is obtained with a minimum of material; yet how this is accomplished is not known. Generations of students have marvelled at the ability of osteoblasts to deposit bone in spiralling lamellar layers around blood vessels, but how it is done remains obscure.

Answers to these questions will undoubtedly benefit from studies of gene disorders causing excessive lack or deposition of bone. The recent identification of loss- and gain-of-function mutations in the osteoblastic receptor LRP5 in families with low and high bone mass is likely to be followed by other exciting discoveries along the same line (see review in this issue by Harada and Rodan, page 349). Coupled with the mapping of quantitative trait loci in large cohorts of individuals, such studies will probably result in identification of components in interacting pathways that regulate skeletal remodeling, as well as in pinpointing sequence variations that contribute to biological variability. The insights to be gained may help improve methods for bone repair, facilitate tissue engineering of skeletal tissues, and allow development of new effective strategies to prevent and cure osteoporosis. □

doi:10.1038/nature01659

- Olsen, B. R., Reginato, A. M. & Wang, W. Bone development. *Annu. Rev. Cell Dev. Biol.* **16**, 191–220 (2000).
- Royce, P. & Steinmann, B. (eds) *Connective Tissue and its Heritable Disorders: Molecular, Genetic, and Medical Aspects* (Wiley, New York, 2002).
- Wilkie, A. O., Oldridge, M., Tang, Z. & Maxson, R. E. Jr Craniosynostosis and related limb anomalies. *Novartis Found. Symp.* **232**, 122–133, 133–143 (2001).
- Hall, C. M. International nosology and classification of constitutional disorders of bone. *Am. J. Med. Genet.* **113**, 65–77 (2002).
- Mundlos, S. & Olsen, B. R. Heritable diseases of the skeleton. Part I: Molecular insights into skeletal development—transcription factors and signaling pathways. *FASEB J.* **11**, 125–132 (1997).

6. Foster, J. W. *et al.* Campomelic dysplasia and autosomal sex reversal caused by mutations in an SRY-related gene. *Nature* **372**, 525–530 (1994).
7. Wagner, T. *et al.* Autosomal sex reversal and campomelic dysplasia are caused by mutations in and around the SRY-related gene SOX9. *Cell* **79**, 1111–1120 (1994).
8. Bi, W. *et al.* Haploinsufficiency of Sox9 results in defective cartilage primordia and premature skeletal mineralization. *Proc. Natl Acad. Sci. USA* **98**, 6698–6703 (2001).
9. Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & de Crombrughe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* **16**, 2813–2828 (2002).
10. Smits, P. *et al.* The transcription factors L-Sox5 and Sox6 are essential for cartilage formation. *Dev. Cell* **1**, 277–290 (2001).
11. Bell, D. M. *et al.* SOX9 directly regulates the type-II collagen gene. *Nature Genet.* **16**, 174–178 (1997).
12. Ducey, P., Zhang, R., Geoffroy, V., Ridall, A. L. & Karsenty, G. Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation. *Cell* **89**, 747–754 (1997).
13. Otto, F. *et al.* Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89**, 765–771 (1997).
14. Komori, T. *et al.* Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* **89**, 755–764 (1997).
15. Mundlos, S. *et al.* Mutations involving the transcription factor CBFA1 cause cleidocranial dysplasia. *Cell* **89**, 773–779 (1997).
16. Kim, I. S., Otto, F., Zabel, B. & Mundlos, S. Regulation of chondrocyte differentiation by Cbfa1. *Mech. Dev.* **80**, 159–170 (1999).
17. Kundu, M. *et al.* Cbfa1 interacts with Runx2 and has a critical role in bone development. *Nature Genet.* **32**, 639–644 (2002).
18. Miller, J. *et al.* The core-binding factor β subunit is required for bone formation and hematopoietic maturation. *Nature Genet.* **32**, 645–649 (2002).
19. Yoshida, C. A. *et al.* Core-binding factor β interacts with Runx2 and is required for skeletal development. *Nature Genet.* **32**, 633–638 (2002).
20. Nakashima, K. *et al.* The novel zinc finger-containing transcription factor Osterix is required for osteoblast differentiation and bone formation. *Cell* **108**, 17–29. (2002).
21. Tassabehji, M. *et al.* Waardenburg's syndrome patients have mutations in the human homologue of the Pax-3 paired box gene. *Nature* **355**, 635–636 (1992).
22. Mansouri, A., Pla, P., Larue, L. & Gruss, P. Pax3 acts cell autonomously in the neural tube and somites by controlling cell surface properties. *Development* **128**, 1995–2005 (2001).
23. Bober, E., Franz, T., Arnold, H. H., Gruss, P. & Tremblay, P. Pax-3 is required for the development of limb muscles: a possible role for the migration of dermomyotomal muscle progenitor cells. *Development* **120**, 603–612 (1994).
24. Epstein, D. J., Vekemans, M. & Gros, P. Sp100, a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3. *Cell* **67**, 767–774 (1991).
25. Hodgkinson, C. A. *et al.* Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**, 395–404 (1993).
26. Tassabehji, M., Newton, V. E. & Read, A. P. Waardenburg syndrome type 2 caused by mutations in the human microphthalmia (MITF) gene. *Nature Genet.* **8**, 251–255 (1994).
27. Steingrimsson, E. *et al.* Molecular basis of mouse microphthalmia (mi) mutations helps explain their developmental and phenotypic consequences. *Nature Genet.* **8**, 256–263 (1994).
28. Weilbaecher, K. N. *et al.* Linkage of M-CSF signaling to Mitf, TFE3, and the osteoclast defect in *Mitf*^{mi/mi} mice. *Mol. Cell* **8**, 749–758 (2001).
29. Steingrimsson, E. *et al.* Mitf and Tfe3, two members of the Mitf-Tfe family of bHLH-Zip transcription factors, have important but functionally redundant roles in osteoclast development. *Proc. Natl Acad. Sci. USA* **99**, 4477–4482 (2002).
30. Motyckova, G. *et al.* Linking osteopetrosis and pycnodysostosis: regulation of cathepsin K expression by the microphthalmia transcription factor family. *Proc. Natl Acad. Sci. USA* **98**, 5798–5803 (2001).
31. Palmeirim, I., Henrique, D., Ish-Horowicz, D. & Pourquie, O. Avian hairy gene expression identifies a molecular clock linked to vertebrate segmentation and somitogenesis. *Cell* **91**, 639–648 (1997).
32. McGrew, M. J., Dale, J. K., Fraboulet, S. & Pourquie, O. The lunatic fringe gene is a target of the molecular clock linked to somite segmentation in avian embryos. *Curr. Biol.* **8**, 979–982 (1998).
33. Bulman, M. P. *et al.* Mutations in the human Delta homologue, *DLL3*, cause axial skeletal defects in spondylocostal dysostosis. *Nature Genet.* **24**, 438–441 (2000).
34. Kusumi, K. *et al.* The mouse pudgy mutation disrupts Delta homologue *Dll3* and initiation of early somite boundaries. *Nature Genet.* **19**, 274–278 (1998).
35. Dunwoodie, S. L. *et al.* Axial skeletal defects caused by mutation in the spondylocostal dysplasia/pudgy gene *Dll3* are associated with disruption of the segmentation clock within the presomitic mesoderm. *Development* **129**, 1795–1806 (2002).
36. Evrard, Y. A., Lun, Y., Aulehla, A., Gan, L. & Johnson, R. L. *lunatic fringe* is an essential mediator of somite segmentation and patterning. *Nature* **394**, 377–381 (1998).
37. Zhang, N. & Gridley, T. Defects in somite formation in *lunatic fringe*-deficient mice. *Nature* **394**, 374–377 (1998).
38. Oda, T. *et al.* Mutations in the human *Jagged1* gene are responsible for Alagille syndrome. *Nature Genet.* **16**, 235–242 (1997).
39. Li, L. *et al.* Alagille syndrome is caused by mutations in human *Jagged1*, which encodes a ligand for Notch1. *Nature Genet.* **16**, 243–251 (1997).
40. Xue, Y. *et al.* Embryonic lethality and vascular defects in mice lacking the Notch ligand Jagged1. *Hum. Mol. Genet.* **8**, 723–730 (1999).
41. McCright, B., Lozier, J. & Gridley, T. A mouse model of Alagille syndrome: *Notch2* as a genetic modifier of *Jag1* haploinsufficiency. *Development* **129**, 1075–1082 (2002).
42. Vortkamp, A., Gessler, M. & Grzeschik, K. H. GLI3 zinc-finger gene interrupted by translocations in Greig syndrome families. *Nature* **352**, 539–540 (1991).
43. Kohlhaase, J., Wischermann, A., Reichenbach, H., Froster, U. & Engel, W. Mutations in the *SALL1* putative transcription factor gene cause Townes-Brocks syndrome. *Nature Genet.* **18**, 81–83 (1998).
44. Muragaki, Y., Mundlos, S., Upton, J. & Olsen, B. R. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* **272**, 548–551 (1996).
45. Hui, C. C. & Joyner, A. L. A mouse model of Greig cephalopolysyndactyly syndrome: the *extra-toes*⁺ mutation contains an intragenic deletion of the *Gli3* gene. *Nature Genet.* **3**, 241–246 (1993).
46. Thien, H. & Ruthner, U. The mouse mutation *Pdn* (Polydactyly Nagoya) is caused by the integration of a retrotransposon into the *Gli3* gene. *Mamm. Genome* **10**, 205–209 (1999).
47. Johnson, K. R. *et al.* A new spontaneous mouse mutation of *Hoxd13* with a polyaniline expansion and phenotype similar to human synpolydactyly. *Hum. Mol. Genet.* **7**, 1033–1038 (1998).
48. McCready, M. E. *et al.* A novel mutation in the *IHH* gene causes brachydactyly type A1: a 95-year-old mystery resolved. *Hum. Genet.* **111**, 368–375 (2002).
49. Oldridge, M. *et al.* Dominant mutations in *ROR2*, encoding an orphan receptor tyrosine kinase, cause brachydactyly type B. *Nature Genet.* **24**, 275–278 (2000).
50. DeChiara, T. M. *et al.* *Ror2*, encoding a receptor-like tyrosine kinase, is required for cartilage and growth plate development. *Nature Genet.* **24**, 271–274 (2000).
51. Afzal, A. R. *et al.* Recessive Robinow syndrome, allelic to dominant brachydactyly type B, is caused by mutation of *ROR2*. *Nature Genet.* **25**, 419–422 (2000).
52. van Bokhoven, H. *et al.* Mutation of the gene encoding the ROR2 tyrosine kinase causes autosomal recessive Robinow syndrome. *Nature Genet.* **25**, 423–426 (2000).
53. Polinkovsky, A. *et al.* Mutations in *CDMP1* cause autosomal dominant brachydactyly type C. *Nature Genet.* **17**, 18–19 (1997).
54. Storm, E. E. *et al.* Limb alterations in *brachypodism* mice due to mutations in a new member of the TGF β -superfamily. *Nature* **368**, 639–643 (1994).
55. Thomas, J. T. *et al.* A human chondrodysplasia due to a mutation in a TGF- β superfamily member. *Nature Genet.* **12**, 315–317 (1996).
56. Gong, Y. *et al.* Heterozygous mutations in the gene encoding noggin affect human joint morphogenesis. *Nature Genet.* **21**, 302–304 (1999).
57. Brunet, L. J., McMahon, J. A., McMahon, A. P. & Harland, R. M. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science* **280**, 1455–1457 (1998).
58. Schipani, E., Kruse, K. & Juppner, H. A constitutively active mutant PTH-PTHrP receptor in Jansen-type metaphyseal chondrodysplasia. *Science* **268**, 98–100 (1995).
59. Jobert, A. S. *et al.* Absence of functional receptors for parathyroid hormone and parathyroid hormone-related peptide in Blomstrand chondrodysplasia. *J. Clin. Invest.* **102**, 34–40 (1998).
60. Hopyan, J. T. *et al.* A mutant PTH/PTHrP type I receptor in enchondromatosis. *Nature Genet.* **30**, 306–310 (2002).
61. Shiang, R. *et al.* Mutations in the transmembrane domain of FGFR3 cause the most common genetic form of dwarfism, achondroplasia. *Cell* **78**, 335–342 (1994).
62. Bellus, G. A. *et al.* A recurrent mutation in the tyrosine kinase domain of fibroblast growth factor receptor 3 causes hypochondroplasia. *Nature Genet.* **10**, 357–359 (1995).
63. Tavorina, P. L. *et al.* Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nature Genet.* **9**, 321–328 (1995).
64. Tavorina, P. L. *et al.* A novel skeletal dysplasia with developmental delay and acanthosis nigricans is caused by a Lys650Met mutation in the fibroblast growth factor receptor 3 gene. *Am. J. Hum. Genet.* **64**, 722–731 (1999).
65. Naski, M. C., Wang, Q., Xu, J. & Ornitz, D. M. Graded activation of fibroblast growth factor receptor 3 by mutations causing achondroplasia and thanatophoric dysplasia. *Nature Genet.* **13**, 233–237 (1996).
66. Ohbayashi, N. *et al.* FGF18 is required for normal cell proliferation and differentiation during osteogenesis and chondrogenesis. *Genes Dev.* **16**, 870–879 (2002).
67. Deng, C., Wynshaw-Boris, A., Zhou, F., Kuo, A. & Leder, P. Fibroblast growth factor receptor 3 is a negative regulator of bone growth. *Cell* **84**, 911–921 (1996).
68. Colvin, J. S., Bohn, B. A., Harding, G. W., McEwen, D. G. & Ornitz, D. M. Skeletal overgrowth and deafness in mice lacking fibroblast growth factor receptor 3. *Nature Genet.* **12**, 390–397 (1996).
69. Minina, E., Kreschel, C., Naski, M. C., Ornitz, D. M. & Vortkamp, A. Interaction of FGF, Ihh/Pthlh, and BMP signaling integrates chondrocyte proliferation and hypertrophic differentiation. *Dev. Cell* **3**, 439–449 (2002).
70. Belin, V. *et al.* *SHOX* mutations in dyschondrosteosis (Leri-Weill syndrome). *Nature Genet.* **19**, 67–69 (1998).
71. Rao, E. *et al.* Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nature Genet.* **16**, 54–63 (1997).
72. Shears, D. J. *et al.* Pseudodominant inheritance of Langer mesomelic dysplasia caused by a *SHOX* homeobox missense mutation. *Am. J. Med. Genet.* **110**, 153–157 (2002).
73. Minns, C. F. *et al.* Histopathological analysis of Leri-Weill dyschondrosteosis: disordered growth plate. *Hand Surg.* **6**, 13–23 (2001).
74. Ross, J. L. *et al.* Mesomelic and rhizomelic short stature: The phenotype of combined Leri-Weill dyschondrosteosis and achondroplasia or hypochondroplasia. *Am. J. Med. Genet.* **116**, 61–65. (2003).

Control of osteoblast function and regulation of bone mass

Shun-ichi Harada & Gideon A. Rodan

Department of Bone Biology & Osteoporosis Research, Merck Research Laboratories, West Point, Pennsylvania 19002, USA
(e-mail: gideon_rodan@merck.com)

The skeleton is an efficient 'servo' (feedback-controlled/steady-state) system that continuously integrates signals and responses which sustain its functions of delivering calcium while maintaining strength. In many individuals, bone mass homeostasis starts failing in midlife, leading to bone loss, osteoporosis and debilitating fractures. Recent advances, spearheaded by genetic information, offer the opportunity to stop or reverse this downhill course.

The functions of bones include maintaining blood calcium levels, providing mechanical support to soft tissues and serving as levers for muscle action, supporting haematopoiesis, and housing the brain and spinal cord. These functions are accomplished by continuous tissue renewal, called remodelling, occurring throughout life at approximately two million microscopic sites in the adult skeleton.

Bone destruction or 'resorption' is carried out by haematopoietically derived osteoclasts. Their number and activity is determined by cell lineage allocation, proliferation and differentiation of osteoclast precursors and the resorptive efficiency of mature osteoclasts. Several effective drugs that control osteoclast generation and/or function are currently available (see review in this issue by Boyle *et al.*, page 337).

Mesenchyme-derived osteoblasts rebuild the resorbed bone by elaborating matrix that then becomes mineralized. Injectable parathyroid hormone (PTH) is the only known agent currently available for pharmacological stimulation of bone formation¹. New insights into osteoblast regulation, reviewed here, could lead to additional bone-building treatments, for example, by controlling the activity of recently discovered genes such *LRP5*, or of newly discovered pathways such as sympathetic neurons. These potential therapies would act in conjunction with all the other factors that control bone mass in humans, some of which are still unknown.

We begin with a brief overview of skeletal homeostasis, and then address a number of questions related to this system, including hierarchies that exist among regulatory factors. This is followed by a review of recent discoveries on the regulation of osteoblast differentiation and function.

Determinants of skeletal homeostasis and bone mass

Bone mass in adults is maintained locally by the balance between osteoclastic bone resorption and osteoblastic bone formation, each of which is subject to controls aimed at fulfilling bone function. Calcium release requires bone destruction, and the principal mediators in this process are PTH and its downstream effector (1,25(OH)₂ vitamin D). Mechanical strength depends on bone mass and its deployment relative to mechanical forces (strain). The local structural adaptation of bones to mechanical loads is the basis for orthopaedic and orthodontic procedures.

Increased mechanical loads stimulate bone formation and suppress resorption, whereas unloading has the opposite effect^{2,3}. The proposed mediators for these effects include direct action on cellular stress-sensitive calcium channels and integrins, as well as paracrine mediators, such as prostaglandin E₂ (PGE₂), prostacyclin and nitric oxide (for a recent review, see ref. 4).

In addition to the important effects on the skeleton of the homeostatic demands for calcium and mechanical adaptation, a third significant input into this 'servo' system is the pronounced influence of sex steroids. Skeletal preservation by oestrogen in females⁵ may be related evolutionarily to the need of calcium stores for embryonic skeletal development. In birds for example, oestrogen causes massive bone formation prior to egg laying. In mammalian adult males and females, including humans, oestrogen inhibits bone resorption by reducing osteoclast number. The mechanism most likely involves lineage allocation of monocyte/macrophage/osteoclast precursors through effects on regulatory cytokines, including interleukin (IL)-1, IL-6, tumour necrosis factor- α and PGE. Oestrogen acts via oestrogen receptor- α in human males and is required for closure of the epiphyses and for reaching or maintaining normal post-pubertal bone mass⁵. The male sex steroid testosterone, responsible for the male phenotype characterized by a larger skeleton, affects bone in several ways. It can be converted by aromatase to oestrogen to inhibit bone resorption, but in addition seems to inhibit bone resorption directly in males and stimulates bone formation in both males and females⁶.

Bone resorption and formation are 'coupled' locally by mechanisms not fully understood, that is, when one goes up or down the other usually follows. But resorption is much faster than formation (it takes at least three months to rebuild bone resorbed in 2–3 weeks). Thus, increased resorption, even when accompanied by coupled increased formation, can cause bone loss owing to these kinetic differences, for example, in oestrogen deficiency or hyperparathyroidism.

Multiple coupling 'factors' attempt to maintain this servo system at its physiological (homeostatic) steady state (Fig. 1)⁷. Bone formation stimulatory factors released from the matrix during resorption, including insulin-like growth factor and transforming growth factor (TGF)- β , may serve this function. Several factors, such as PTH, PGE, fibroblast growth factor, TGF- β and even RANK ligand, have been

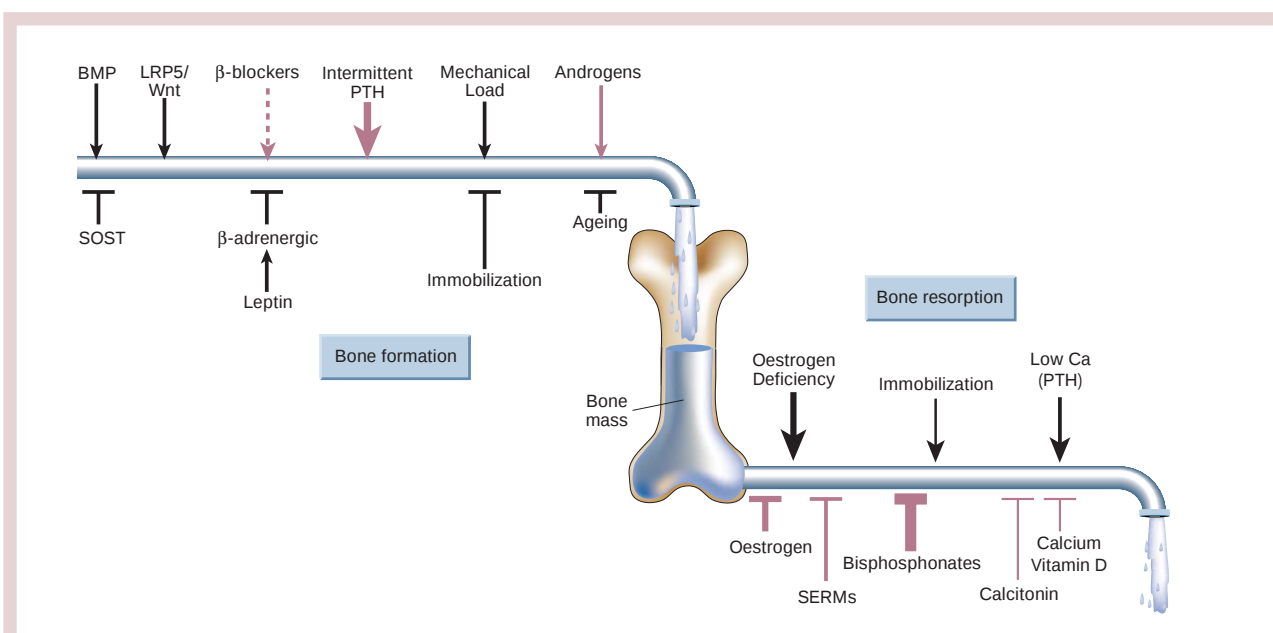


Figure 1 Determinants of skeletal homeostasis and bone mass. Schematic representation of the servo system that maintains bone mass at steady-state levels. Physiological (blue) and pharmacological (orange) stimulators and inhibitors of bone formation and resorption are listed. The relative impact, where known, is represented

by the thickness of the arrows. Solid lines are current therapies and dotted lines putative ones. Abbreviations: BMP, bone morphogenetic protein(s); SOST, sclerostin; LRP5, low-density lipoprotein (LDL)-receptor-related protein 5; PTH, parathyroid hormone; SERM, selective oestrogen-receptor modulator.

shown to stimulate both resorption and formation. Mechanical effects on bone could also couple resorption to formation, as weakening of the bone as a result of resorption should engender corrective bone formation⁶.

Steady state in servo systems is reached and maintained by feedback loops. Here, the best characterized is calcium control of PTH secretion, which in turn regulates circulating calcium levels via bone resorption, intestinal absorption and renal reabsorption. Sex hormone levels are regulated by feedback to the pituitary, but are controlled by the reproductive agenda, rather than skeletal needs. Mechanical feedback signals, less well defined at the molecular level, must emanate from the strain in the bone matrix, instructing osteoclasts and osteoblasts to increase, decrease or stop their activity⁴.

This homeostatic system, like all others, is subject to genetic influences. This is reflected in the tight correlation of bone mineral density between identical twins compared with non-identical twins, and the lower bone density in daughters of osteoporotic mothers^{9,10}. Multiple genes are undoubtedly involved, a conclusion supported by studies in inbred mice, which have identified several quantitative trait loci responsible for strain dependence of bone mass¹¹.

Hierarchy among factors regulating bone mass

In-depth discussion of skeletal homeostasis is beyond the scope of this review, but the following three questions will help to place recent advances in a pathophysiological and therapeutic context. First, is there a hierarchy in bone functions; second, is there a hierarchy among the missing factor(s) responsible for bone loss and osteoporosis; and third, which external (pharmacological) inputs can compensate and reverse the bone loss?

Concerning the first question, calcium mobilization overrides other functions of the skeleton. Calcium deficiency due to renal disease, malabsorption, other pathology or poor calcium diet invariably causes bone loss, mediated by elevated PTH. But owing to other servo inputs (for example, mechanical function), this loss is not random throughout the skeleton, the bone most exposed to mechanical loads being most protected¹².

The effect of oestrogen seems to trump mechanical function, as exercise is limited in its ability to maintain or restore bone mass in postmenopausal women¹³ and amenorrhoeic marathon runners lose bone. Among the three modulators of bone mass — calcium availability, sex steroids and mechanical usage — the last has the least pronounced effects. Excessive reductions in bone strain produced by weightlessness (microgravity in outer space) or immobilization (paralysis, prolonged bed rest or application of casts) can cause significant bone loss, while strenuous athletic activity (for example, professional tennis) can augment certain bones¹⁴. However, physiological changes in physical function, such as sedentary lifestyle or moderate exercise, have only a slow and modest effect on bone mass¹³.

Regarding factors responsible for osteoporosis, from an epidemiological perspective oestrogen is clearly at the top of the list. But other inputs into this servo system (such as dietary calcium, parathyroid function and genetic background) must also be important — although their contribution has not been fully quantified — as not all postmenopausal women develop osteoporosis. Additionally, bone loss can occur before menopause in women or the onset of testosterone reduction in men.

The recent discoveries discussed below could be among the inputs that may prevent or reverse bone loss, while future studies will show if and how they relate to the main skeletal modulators (calcium regulation, biomechanics and sex steroids) and what role they may have in the pathophysiology of osteoporosis. This leads to the third question, the ability of various therapies to prevent or reverse bone loss. In principle, replacement of a missing factor(s) should correct its deficiency (for example, oestrogen replacement therapy or ERT). However, in servo systems, inputs not directly related to the deficiency can move the steady-state level in the desired direction. For example, glucocorticoid-induced osteoporosis, caused at least in part by decreased bone formation, can be treated effectively with inhibitors of bone resorption (bisphosphonates). Intermittent PTH administered to stimulate bone formation does not replace PTH deficiency, but is an effective pharmacological strategy that takes advantage of one of the actions of this hormone. Any input that

can move the system in the desired direction can, in principle, have therapeutic utility. Its efficacy will depend on its role in the system in the context of the other inputs. One important feature of a servo system is that when it is brought to a new steady state by a pharmacological intervention, cessation of treatment will bring it back to the previous level (for example, bone loss following cessation of ERT or intermittent PTH).

With this background in mind we shall briefly review several recent discoveries, starting with central nervous system (CNS) control of bone mass, which came to light through observations in mice with genetic abnormalities in hypothalamic regulation of body weight and reproduction.

Central control of bone formation

The concept that the CNS may control bone formation might have been hinted at by the surge in osteogenesis following head injury, but was first proposed by Ducy and co-workers in 2000 as an explanation for the inhibitory effect of leptin on bone formation¹⁵. The study was prompted by the rapid recovery of bone mass following severe osteopenia (decreased bone mass) caused by inducible osteoblast ablation, which suggested precise control of bone mass homeostasis¹⁶.

Another piece of the puzzle was provided by the surprising high bone mass (HBM) phenotype in leptin-deficient *ob/ob* mice that have low sex steroid and high corticosteroid concentrations¹⁵. Leptin, produced by fat tissue, acts in the hypothalamus to regulate body weight and fat mass through appetite suppression and increased energy expenditure. HBM, associated with increased bone formation, was also observed in leptin receptor-deficient mice as well as in leptin-deficient lean lipodystrophic mice, pointing to a role for leptin signalling per se, rather than obesity, in the bone phenotype. Moreover, no similar findings were observed in mice with mutations in the α -melanocyte-stimulating hormone (α -MSH)/agouti/melanocortin-4 (MC-4)-receptor signalling pathway or the cocaine- and amphetamine-regulated transcript (CART) pathway, all of which are involved in the anorexigenic action of leptin¹⁷ (Fig. 2), indicating that the anti-osteogenic action of leptin can be dissociated from its anti-appetite effects. These authors found no direct effects of leptin on bone cells; leptin suppressed bone formation and reduced bone mass, but only when injected intracerebroventricularly (ICV). The failure to document humoral factor mediation, using extensive pharmacophysiological analyses (including a parabiosis study¹⁷), pointed to central control of bone metabolism. Other investigators reported the expression of leptin receptors in bone cells and stimulatory effects on bone^{18,19}, leaving open the possibility of dual action (central and peripheral) of leptin²⁰.

Noting the low bone mass phenotype in dopamine transporter-deficient mice, which lack dopamine uptake into presynaptic terminals²¹, Takeda *et al.*¹⁷ tested the possible involvement of the sympathetic nervous system (SNS), which is known to be activated by leptin signalling. Mice deficient in dopamine β -hydroxylase, which is necessary for production of noradrenaline and adrenaline, had HBM that did not respond to ICV infusion of leptin. In addition, bone formation and bone mass decreased after treatment with the β -adrenergic agonist isoproterenol and increased with propranolol, a β -adrenergic antagonist. These responses were resistant to leptin infusion, leading to the conclusion that SNS is the downstream mediator of leptin's central control of bone formation¹⁷ (Fig. 2).

Central control was also supported by increased bone formation and bone mass following inactivation of the Y2 neuropeptide Y (NPY) receptor in brain²². Although no bone phenotype has been reported in NPY-null mice, this observation is consistent with the decrease in bone formation caused by ICV injection of NPY, suggesting the central control of bone formation by Y2 and its ligands¹⁵. Because leptin suppresses NPY, this pathway does not mediate the effects of leptin on bone. Nevertheless, it raises the

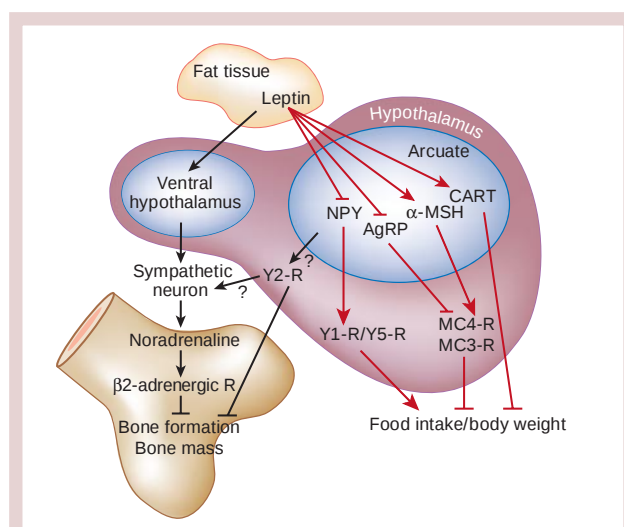


Figure 2 Leptin signalling pathways. In the regulation of body weight, leptin acts on the arcuate nucleus to induce the anorexigenic peptides α -melanocyte-stimulating hormone (α -MSH) and cocaine- and amphetamine-related transcript (CART), and to repress the orexigenic peptides neuropeptide Y (NPY) and agouti-related protein (AgRP), an antagonist for the melanocortin-3 (MC-3) and MC-4 receptors. These combined effects of leptin lead to increased body weight mainly via MC-3 and MC-4 receptors and Y1 and Y5 NPY receptors. Leptin also increases energy expenditure by activating sympathetic neurons. In contrast to leptin's anorexigenic action, during bone formation the anti-osteogenic action of leptin is mediated by the sympathetic nervous system (SNS) via the ventral hypothalamus. Noradrenaline released through activation of the SNS acts on β 2-adrenergic receptors expressed in osteoblasts to inhibit bone formation. Activation of Y2 NPY receptors in the hypothalamus by its ligands also represses bone formation.

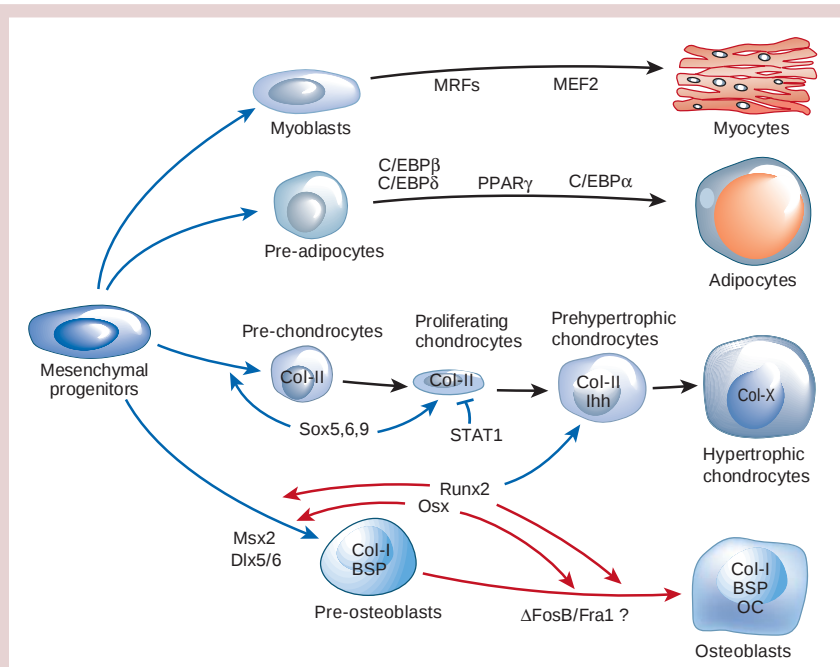
question of whether leptin and Y2 act on bone through a common pathway. While Y1 and Y5 receptors mediate orexigenic effects of NPY²³, pharmacological data suggest that the Y2 receptor could be involved in NPY regulation of sympathetic neuronal activity^{24,25}. But NPY action on the autonomic nervous system is complex, involving both peripheral and central pathways, and additional studies are needed to evaluate whether the effects of NPY on bone and the SNS are linked.

These exciting discoveries raise a number of questions that future studies should address. The mediator of the central anti-osteogenic action of leptin in bone is the β 2-adrenergic receptor. This receptor is known to activate the cAMP signalling pathway, which has been implicated in the bone formation stimulatory effect of PTH and prostaglandins. These pro- and antianabolic effects could be mediated by different target cells or cAMP may have different effects based on the pattern of its changes.

Another question pertains to the role of CNS and SNS in the hierarchy of physiological and pathological regulators of bone mass. Does this central control represent the master regulator of bone formation that, for example, is important in restoring bone mass after osteoblast ablation?

Last, do the CNS and SNS, and for that matter leptin, have similar roles in the regulation of bone formation in humans as they do in mice. For example, bone loss observed in anorexia nervosa patients or amenorrhoeic female athletes, both with low levels of leptin and oestrogens, suggests that, unlike in mice, leptin deficiency cannot override bone loss induced by oestrogen deficiency in humans²⁶. Stimulatory effects of propranolol on fracture repair have been reported in rats²⁷, but limited information has been available on the role for SNS in the human skeleton. Supportive evidence is offered by the association of osteoporosis with 'reflex sympathetic dystrophy'. This condition is caused by hyperadrenergic activity and is treated

Figure 3 Transcriptional control of osteoblastic, chondrocytic, adipocytic and myocytic differentiation. Osteoblasts differentiate from mesenchymal progenitor cells that also give rise to myocytes, under the control of MRFs and MEF2³¹, to adipocytes under the control of C/EBP α , β and δ and PPAR γ ³⁰, and to chondrocytes under the control of Sox5, -6 and -9³³ and STAT1 (see review in this issue by Kronenberg, page 332). Runx2 is essential for osteoblast differentiation and is also involved in chondrocyte maturation. Osterix (Osx) acts downstream of Runx2 to induce mature osteoblasts that express osteoblast markers, including osteocalcin. Abbreviations: MRFs, myogenic regulatory factors (including MyoD, myogenin, myogenic factor 5 and myogenic regulatory factor 4); MEF2, myocyte-enhancer factor 2; C/EBP, CCAAT-enhancer-binding protein; PPAR γ , peroxisome proliferator-activated receptor γ ; STAT1, signal transducers and activators of transcription-1; Runx2, runt-related transcription factor 2; Col-I/II/X, type I/II/X collagen; Ihh, Indian hedgehog; BSP, bone sialoprotein; OC, osteocalcin.



with β -blockers²⁸. As β -blockers have been widely used in the clinic, insights into their effect on the skeleton could be obtained retrospectively.

Transcriptional regulation of bone formation

Genetic studies in mice have also provided new insights into the transcriptional regulation of osteoblast differentiation. Osteoblasts are derived from mesenchymal precursor cells that also give rise to chondrocytes, myoblasts, adipocytes and tendon cells²⁹. Transcriptional control of myogenesis and adipogenesis consists of a cascade of transcriptional events driven by a series of transcription factors that control each other's phenotype-specific gene expression^{30,31} (Fig. 2). Understanding the transcriptional control of osteoblast/chondrocyte differentiation has been more limited, largely owing to the lack of cell-culture systems that fully recapitulate osteoblast/chondrocyte differentiation in a synchronized fashion. Based on skeletal phenotype associated with genetic mutations in animals and humans and the analysis of transcriptional complexes formed with osteoblast/chondrocyte-specific enhancers, a number of transcription factors that control osteogenesis and chondrogenesis have been identified³². These include runt-related transcription factor 2 (Runx2) and sex determining region Y-box 9 (Sox9), 'master' regulators of osteogenesis and chondrogenesis^{33–36} (see review in this issue by Kronenberg, page 332).

The essential role of Runx2, also known as *Cbfa1*, *Osf2* and *AML3*, in osteoblast differentiation was documented unambiguously in null-mutation mice that had a cartilaginous skeleton with complete absence of osteoblasts^{34–36}. Heterozygous mice had a defect in intramembranous ossification that resembles cleidocranial dysplasia in humans, caused by mutation of *RUNX2* (ref. 36; and see review in this issue by Zelzer and Olsen, page 343). Although cartilage develops in Runx2-null mice, careful histological analysis showed delayed chondrocyte maturation³⁷, consistent with Runx2 expression in hypertrophic chondrocytes. Moreover, transgenic expression of Runx2 in chondrocytes, via the chondrocyte-specific type II collagen promoter, results in ectopic chondrocyte hypertrophy and endochondral ossification^{38,39}, indicating that Runx2 controls differentiation of hypertrophic chondrocytes and osteoblasts.

This dual role for Runx2 suggested the presence of additional factors that act in osteoblasts to control osteogenesis. One such factor

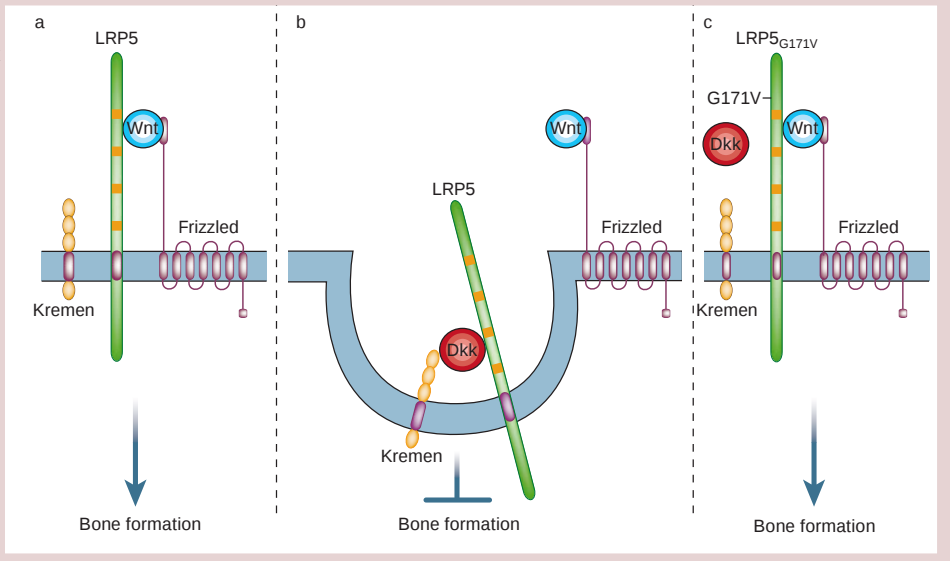
is *Osterix* (*Osx*)⁴⁰. *Osx* is a zinc finger-containing protein induced in C2C12 myoblasts in response to bone morphogenetic protein (BMP), a potent stimulator of bone formation when injected into muscle or dermis. *Osx*-null mice develop a perfectly patterned skeleton composed entirely of cartilage, lacking osteoblasts and mineralized bone matrix. Unlike Runx2-null mice, the cartilage of *Osx*-null mice is normal, containing fully mineralized, terminally differentiated hypertrophic chondrocytes, pointing to a specific role for *Osx* in osteoblast differentiation. *Osx* is not expressed in Runx2-null mice, while the expression of Runx2 is normal in *Osx*-null mice. Interestingly, *Osx*-null osteoblast precursors in the periosteum of membranous bones express chondrocyte markers, such as *Sox9* and *Col2a1*, suggesting that *Osx* acts downstream of Runx2 to induce osteoblastic differentiation in bipotential chondro-osteoprogenitor cells.

A number of transcription factors/cofactors have been shown to interact with Runx2. AJ18, a zinc finger-containing factor, inhibits Runx activity by competing for its DNA-binding sequence⁴¹. Core-binding factor- β (*Cbfb*) is a heterodimerizing partner of Runx1 and Runx3 that is essential for haematopoiesis. Transgenic rescue of embryonic lethal *Cbfb*-null mice and 'knock-in' of *Cbfb* fused in-frame to a cDNA encoding green fluorescent protein^{42,43} resulted in mice that exhibited delayed ossification, indicating a role for *Cbfb* in bone. However, unlike Runx2-null mice that completely lack bone and osteoblasts, ossification is initiated in these mice, suggesting that Runx2 can act in the absence of *Cbfb*.

Distal-less homeobox 5 (*Dlx5*) and *msh homeobox homologue 2* (*Msx2*) are homeobox-containing transcription factors expressed in early stages of osteoblast differentiation. Functional analysis *in vitro* points to reciprocal roles of *Dlx5* and *Msx2* as activator and repressor of transcription. Murine and human mutations show that they are essential for normal intramembranous ossification⁴⁴. *Dlx5/Dlx6*-null mice also indicate a role for *Dlx5/Dlx6* in the axial and appendicular skeleton⁴⁵.

Changes in Fos-family transcription factors affect both osteoblasts and osteoclasts⁴⁶. Transgenic expression of *cfos* in mice causes osteosarcoma, whereas *cfos*-null mutation causes osteopetrosis. Transgenic expression of *fra-1* or Δ FosB (an alternatively spliced form of FosB), both widely expressed in many tissues in addition to bone, stimulates bone formation and increases bone mass in

Figure 4 A model for LRP5 signalling pathway in bone. **a**, Wnt induces the complex of LRP5 (a co-receptor for Wnt) and Frizzled (the receptor for Wnt), which activates the canonical Wnt signalling pathway, leading to bone formation. **b**, LRP5 can also form a ternary complex with Dickkopf-1 (Dkk; a Wnt inhibitor) and Kremen (a receptor of Dkk), which triggers rapid internalization and depletion of cell-surface LRP5, leading to inhibition of the canonical Wnt signalling pathway⁶⁷. **c**, LRP5_{G171V} — the mutant LRP5 that causes high bone mass — probably forms complexes preferentially with Wnt and Frizzled in bone tissue, leading to constitutive activation of Wnt signalling and bone formation (hypothetical model). Figure modified from ref. 67.



mice^{47,48}. It is not known how these Fos-family proteins, which lack transactivation domains, increase bone formation.

Recent insights into the transcriptional regulation of osteoblast differentiation during development are largely due to advances in human and mouse genetics (Fig. 3). But information on the transcriptional regulation of bone formation and bone mass in the adult, which could be most useful for therapeutic applications, remains limited. Transgenic expression of a dominant negative form of *Runx2* in mature osteoblasts, under the mouse osteoblast-specific osteocalcin promoter, decreases bone formation and reduces bone mass, supporting a role for *Runx2* in osteoblast activity in adult mice⁴⁹. Transgenic expression of *Runx2* under the type I collagen promoter, an early marker for osteoblast commitment, results in a surprising osteopenic phenotype caused by impaired osteoblast maturation and increased bone resorption^{50,51}. These results suggest a potential role for *Runx2* in bone resorption, consistent with the absence of osteoclasts in *Runx2*-null mice, and support the notion that temporal control of *Runx2* activity is important for normal bone development. Conditional inactivation of genes, in a spatial and temporal fashion, should clarify the fine aspects of transcriptional regulation of bone cell differentiation and activity.

New genes responsible for high bone mass in humans

New genes responsible for hereditary skeletal disorders could provide unexpected therapeutic opportunities. These include mutations associated with increased bone formation manifested as HBM. The affected individuals are generally healthy, suggesting that these genes have tissue-specific roles. Two such genes are *LRP5* and *SOST*.

LRP5 encodes the low-density lipoprotein (LDL)-receptor-related protein 5, responsible for HBM in one kindred with strikingly dense bone and no other abnormalities. This trait was mapped to chromosome 11q13 (the long arm of chromosome 11 at band 13) where a single mutation, G171V, was found in all affected individuals⁵². The identical mutation was identified in another HBM kindred⁵³. *LRP5* was also mapped as the locus for osteoporosis-pseudoglioma syndrome (OPPG), an autosomal recessive disease characterized by severe osteoporosis due to decreased bone formation and pseudoglioma resulting from failed regression of primary vitreal vasculature⁵⁴. Thus, while gain of function leads to HBM, an autosomal dominant trait, loss of function leads to osteoporosis.

The role of *Lrp5* in bone formation and bone mass was further confirmed genetically in mice. Null mutation of *Lrp5* results in post-natal bone loss, due to decreased bone formation and osteoblast

proliferation, in a *Runx2*-independent manner⁵⁵. Conversely, increased bone formation and higher bone mass was reported in transgenic mice that express in osteoblasts *LRP5* with the HBM mutation G171V (reported by F. Bex and co-workers at the 2002 meeting of the American Society for Bone and Mineral Research), supporting a cell-autonomous role for *Lrp5* in osteoblasts. No developmental abnormalities were observed in the skeleton of these mice, consistent with a unique role of *Lrp5* in bone homeostasis.

The identification of *LRP5* as a key molecule in bone regulation was surprising for several reasons. *LRP5* is expressed at low levels in almost all tissues and shows little temporal changes. Second, the *LRP5* signal transduction pathways were not known to control bone formation. *LRP5* was identified originally as a member of the LDL-receptor superfamily with low affinity to apolipoprotein E⁵⁶. However, recent *in vitro* findings support a role for *LRP5* in Wntless/Wnt signalling, largely because of its close homology to *Lrp6*, a mammalian homologue of Arrow, which is a co-receptor for Wntless signalling in *Drosophila*. Null mutation of *Lrp6* in mice results in loss of Wnt signalling, which is important for many developmental processes⁵⁷. Consistent with its structural homology to *LRP6*, *LRP5* interacts *in vitro* in cell culture with a subfamily of Wnts, including Wnt1 and Wnt3a, to form a complex with Frizzled, a well characterized receptor for Wnts, leading to activation of the canonical Wnt signalling pathway^{55,58} (Fig. 4a). Although Wnt/Wingless signalling has been linked to limb development and chondrogenesis, there was little information on its role in bone and osteoblasts. Activation of Wnt signalling in osteoblast precursor cells by Wnt3a, a constitutively active form of β -catenin or LiCl (an activator of Wnt signalling), was shown recently to promote osteoblastic differentiation^{54,59}.

Involvement of the Wnt pathway in the action of *LRP5* on bone is supported by the observation that *LRP5* with the HBM mutation prevents inhibition of Wnt signalling by Dickkopf-1^{53,60} (Fig. 4b, c). And recent subtractive hybridization and microarray analysis identified a series of genes involved in Wnt signalling as signature genes induced during fracture repair in rodents⁶¹. Taken together, these genetic and limited biological observations support a previously unrecognized role for *LRP5* and Wnt signalling in bone formation (Fig. 4). It remains to be shown how *LRP5* plays such a selective role in bone. Six additional mutations of the *LRP5* gene in the amino-terminal domain near G171 were identified recently in individuals with increased bone density, notably in cortical bone⁶². This exciting discovery points to the power of genetics in the identification of new mechanisms that could not have been hypothesized based on previous knowledge.

The recent discovery of the gene *Sclerostin* (*SOST*) in chromosome 17q12–q21, provided the first evidence for a role of BMP in bone mass regulation. *SOST* is the locus for sclerosteosis, an autosomal recessive HBM trait initially found in South African Dutch descendants with elevated bone formation. Five distinct mutations that lead to *SOST* inactivation have been identified in families with HBM in South Africa, Senegal, Brazil and the United States^{63,64}. *SOST* encodes a protein homologous to the secreted factors that regulate BMPs, such as Noggin, Chordin and Gremlin⁶⁵. Unlike other BMP-binding proteins that are expressed in many tissues, *SOST* expression is restricted to bone and cartilage, except for low expression in liver and kidney⁶⁴. Preliminary data suggest that *SOST* decreases bone formation by suppressing BMP activity in bone. This is the first evidence for a potential role of BMP in adult bone since its discovery 30 years ago as a potent inducer of ectopic osteogenesis.

From the bench to the clinic

The spectacular advances in genomics and genetics promised to usher in an unprecedented era of understanding and treating disease. Three years after deciphering the human genome it is clear that the progress is slow, for several reasons. Many common diseases, such as osteoporosis, are multigenic and result from the interplay between genetic and environmental factors. Even when disease rate-limiting genes are identified, they have to be amenable to drug discovery — that is, available pharmaceutical tools should be able to mimic or compensate for their absence (lack of function) or excessive activity (gain of function). Another requirement is sufficient tissue selectivity, to avoid mechanism-based ‘side effects’. There are, however, examples in bone diseases where these requirements could be met⁶⁶.

In bone resorption, genetic approaches have identified osteoprotegerin, cathepsin K and the chloride channel 7 (CLCN7) as rate limiting for osteoclast generation and activity, and all three are now drug discovery targets (see review in this issue by Boyle and co-workers, page 337). Some of the genetic discoveries described above could lead to treatments that increase bone formation. Therapeutic modulation of transcription factors, except for nuclear receptors controlled by cognate ligands, has been historically difficult, as it occurs beyond the cellular and nuclear membranes and involves multiple large interactive proteins. Thus, exploiting for therapeutic purposes the information provided by *RunX2*, *Osx*, *Dlx1/Msx* and *Fos*, for example, faces significant challenges, which could possibly be met on a case by case basis.

There may be better prospects in using the *SOST* discovery for therapeutic purposes, as *SOST* interaction with BMP most likely occurs extracellularly. One could conceive that a blocking anti-*SOST* antibody could interact with the binding domain and mimic its deletion.

The LRP5 mutation was suggested to increase skeletal responsiveness to mechanical stimulation. At the biochemical level it apparently increases Wnt signalling in bone. As mentioned above, one needs to understand the tissue specificity of this effect and try to mimic it pharmacologically. The challenges in this case are the ubiquitous expression of LRP5 and involvement of Wnt in multiple processes, including cancer.

Central regulation of bone mass via the SNS offers a new dimension for potential bone therapies. There is currently too little information to speculate if central or peripheral approaches to activate this pathway, if applicable to humans, are more promising. As with other potential therapies, achieving tissue selectivity will be the main challenge here. However, each of these three potential therapies involving *SOST*, LRP5 and SNS deal with modulation of receptor–ligand interactions that have in many cases been amenable to successful therapeutic approaches. □

doi:10.1038/nature01660

- Neer, R. M. *et al.* Effect of parathyroid hormone (1–34) on fractures and bone mineral density in postmenopausal women with osteoporosis. *N. Engl. J. Med.* **344**, 1434–1441 (2001).
- Rodan, G. A. Bone mass homeostasis and bisphosphonate action. *Bone* **20**, 1–4 (1997).
- Ehrlich, P. J. & Lanyon, L. E. Mechanical strain and bone cell function: a review. *Osteoporosis Int.* **13**, 688–700 (2002).

- Pavalko, F. M. *et al.* A Model for mechanotransduction in bone cells: The load-bearing mechanosomes. *J. Cell. Biochem.* **88**, 104–112 (2003).
- Riggs, B. L., Khosla, S. & Melton, L. J. III Sex steroids and the construction and conservation of the adult skeleton. *Endocr. Rev.* **23**, 279–302 (2002).
- Schmidt, A., Harada, S. & Rodan, G. A. in *Principles of Bone Biology* (eds Bilezikian, J. P., Raisz, L. G. & Rodan, G. A.) 1455–1466 (Academic, San Diego, 2002).
- Frost, H. M. Cybernetic aspects of bone modeling and remodeling with special reference to osteoporosis and whole-bone strength. *Am. J. Hum. Biol.* **13**, 235–248 (2001).
- Martin, T. J. & Rodan, G. A. in *Osteoporosis* (eds Marcus, R., Feldman, D. & Kelsey, J.) 361–371 (Academic, San Diego, 2002).
- Pocock, N. A. *et al.* Genetic determinants of bone mass in adults. A twin study. *J. Clin. Invest.* **80**, 706–710 (1987).
- Seeman, E. *et al.* Reduced bone mass in daughters of women with osteoporosis. *N. Engl. J. Med.* **320**, 554–558 (1989).
- Ralston, S. H. Genetic control of susceptibility to osteoporosis. *J. Clin. Endocrinol. Metab.* **87**, 2460–2466 (2002).
- Mosekilde, L. Consequences of the remodelling process for vertebral trabecular bone structure: a scanning electron microscopy study (uncoupling of unloaded structures). *Bone Miner.* **10**, 13–35 (1990).
- Wallace, B. A. & Cumming, R. G. Systematic review of randomized trials of the effect of exercise on bone mass in pre- and postmenopausal women. *Calcif. Tissue Int.* **67**, 10–18 (2000).
- Kontulainen, S. *et al.* Good maintenance of exercise-induced bone gain with decreased training of female tennis and squash players: a prospective 5-year follow-up study of young and old starters and controls. *J. Bone Miner. Res.* **16**, 195–201 (2001).
- Ducy, P. *et al.* Leptin inhibits bone formation through a hypothalamic relay: a central control of bone mass. *Cell* **100**, 197–207 (2000).
- Corral, D. A. *et al.* Dissociation between bone resorption and bone formation in osteopenic transgenic mice. *Proc. Natl Acad. Sci. USA* **95**, 13835–13840 (1998).
- Takeda, S. *et al.* Leptin regulates bone formation via the sympathetic nervous system. *Cell* **111**, 305–317 (2002).
- Thomas, T. *et al.* Leptin acts on human marrow stromal cells to enhance differentiation to osteoblasts and to inhibit differentiation to adipocytes. *Endocrinology* **140**, 1630–1638 (1999).
- Burguera, B. *et al.* Leptin reduces ovariectomy-induced bone loss in rats. *Endocrinology* **142**, 3546–3553 (2001).
- Khosla, S. Leptin—central or peripheral to the regulation of bone metabolism? *Endocrinology* **143**, 4161–4164 (2002).
- Blizotes, M. *et al.* Bone histomorphometric and biomechanical abnormalities in mice homozygous for deletion of the dopamine transporter gene. *Bone* **26**, 15–19 (2000).
- Ballock, P. A. *et al.* Hypothalamic Y2 receptors regulate bone formation. *J. Clin. Invest.* **109**, 915–921 (2002).
- Thorsell, A. & Heilig, M. Diverse functions of neuropeptide Y revealed using genetically modified animals. *Neuropeptides* **36**, 182–193 (2002).
- Mahns, D. A., Lacroix, J. S. & Potter, E. K. Inhibition of vagal vasodilation by a selective neuropeptide Y Y2 receptor agonist in the bronchial circulation of anaesthetized dogs. *J. Auton. Nerv. Syst.* **73**, 80–85 (1998).
- Smith-White, M. A., Herzog, H. & Potter, E. K. Role of neuropeptide Y Y2 receptors in modulation of cardiac parasympathetic neurotransmission. *Regul. Pept.* **103**, 105–111 (2002).
- Flier, J. S. Physiology: is brain sympathetic to bone? *Nature* **420**, 619–622 (2002).
- Minkowitz, B., Boskey, A. L., Lane, J. M., Pearlman, H. S. & Vigorita, V. J. Effects of propranolol on bone metabolism in the rat. *J. Orthop. Res.* **9**, 869–875 (1991).
- Schwartzman, R. J. New treatments for reflex sympathetic dystrophy. *N. Engl. J. Med.* **343**, 654–656 (2000).
- Ducy, P., Schinke, T. & Karsenty, G. The osteoblast: a sophisticated fibroblast under central surveillance. *Science* **289**, 1501–1504 (2000).
- Rosen, E. D. & Spiegelman, B. M. Molecular regulation of adipogenesis. *Annu. Rev. Cell Dev. Biol.* **16**, 145–171 (2000).
- Arnold, H. H. & Winter, B. Muscle differentiation: more complexity to the network of myogenic regulators. *Curr. Opin. Genet. Dev.* **8**, 539–544 (1998).
- Karsenty, G. & Wagner, E. F. Reaching a genetic and molecular understanding of skeletal development. *Dev. Cell* **2**, 389–406 (2002).
- Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A. & de Crombrughe, B. The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of *Sox5* and *Sox6*. *Genes Dev.* **16**, 2813–2828 (2002).
- Ducy, P., Zhang, R., Geoffroy, V., Ridall, A. L. & Karsenty, G. *Osf2/Cbfa1*: a transcriptional activator of osteoblast differentiation. *Cell* **89**, 747–754 (1997).
- Komori, T. *et al.* Targeted disruption of *Cbfa1* results in a complete lack of bone formation owing to maturational arrest of osteoblasts. *Cell* **89**, 755–764 (1997).
- Otto, F. *et al.* *Cbfa1*, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development. *Cell* **89**, 765–771 (1997).
- Inada, M. *et al.* Maturation disturbance of chondrocytes in *Cbfa1*-deficient mice. *Dev. Dyn.* **214**, 279–290 (1999).
- Ueta, C. *et al.* Skeletal malformations caused by overexpression of *Cbfa1* or its dominant negative form in chondrocytes. *J. Cell Biol.* **153**, 87–100 (2001).
- Takeda, S., Bonnamy, J. P., Owen, M. J., Ducy, P. & Karsenty, G. Continuous expression of *Cbfa1* in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues *Cbfa1*-deficient mice. *Genes Dev.* **15**, 467–481 (2001).
- Nakashima, K. *et al.* The novel zinc finger-containing transcription factor osterix is required for osteoblast differentiation and bone formation. *Cell* **108**, 17–29 (2002).
- Jheon, A. H., Ganss, B., Cheifetz, S. & Sodek, J. Characterization of a novel KRAB/C2H2 zinc finger transcription factor involved in bone development. *J. Biol. Chem.* **276**, 18282–18289 (2001).
- Kundu, M. *et al.* *Cbfb* interacts with *Runx2* and has a critical role in bone development. *Nature Genet.* **32**, 639–644 (2002).
- Yoshida, C. A. *et al.* Core-binding factor β interacts with *Runx2* and is required for skeletal development. *Nature Genet.* **32**, 633–638 (2002).
- Bendall, A. J. & Abate-Shen, C. Roles for *Msx* and *Dlx* homeoproteins in vertebrate development. *Gene* **247**, 17–31 (2000).

45. Robledo, R. F., Rajan, L., Li, X. & Lufkin, T. The *Dlx5* and *Dlx6* homeobox genes are essential for craniofacial, axial, and appendicular skeletal development. *Genes Dev.* **16**, 1089–1101 (2002).
46. Grigoriadis, A. E., Wang, Z. Q. & Wagner, E. F. *Fos* and bone cell development: lessons from a nuclear oncogene. *Trends Genet.* **11**, 436–441 (1995).
47. Jochum, W. *et al.* Increased bone formation and osteosclerosis in mice overexpressing the transcription factor Fra-1. *Nature Med.* **6**, 980–984 (2000).
48. Sabatakos, G. *et al.* Overexpression of Δ FosB transcription factor(s) increases bone formation and inhibits adipogenesis. *Nature Med.* **6**, 985–990 (2000).
49. Ducy, P. *et al.* A *Cbfa1*-dependent genetic pathway controls bone formation beyond embryonic development. *Genes Dev.* **13**, 1025–1036 (1999).
50. Geoffroy, V., Kneissel, M., Fournier, B., Boyde, A. & Matthias, P. High bone resorption in adult aging transgenic mice overexpressing *Cbfa1/Runx2* in cells of the osteoblastic lineage. *Mol. Cell. Biol.* **22**, 6222–6233 (2002).
51. Liu, W. *et al.* Overexpression of *Cbfa1* in osteoblasts inhibits osteoblast maturation and causes osteopenia with multiple fractures. *J. Cell Biol.* **155**, 157–166 (2001).
52. Little, R. D. *et al.* A mutation in the LDL receptor-related protein 5 gene results in the autosomal dominant high-bone-mass trait. *Am. J. Hum. Genet.* **70**, 11–19 (2002).
53. Boyden, L. M. *et al.* High bone density due to a mutation in LDL-receptor-related protein 5. *N. Engl. J. Med.* **346**, 1513–1521 (2002).
54. Gong, Y. *et al.* LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell* **107**, 513–523 (2001).
55. Kato, M. *et al.* *Cbfa1*-independent decrease in osteoblast proliferation, osteopenia, and persistent embryonic eye vascularization in mice deficient in *Lrp5*, a Wnt coreceptor. *J. Cell Biol.* **157**, 303–314 (2002).
56. Hey, P. J. *et al.* Cloning of a novel member of the low-density lipoprotein receptor family. *Gene* **216**, 103–111 (1998).
57. Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. & Skarnes, W. C. An LDL-receptor-related protein mediates Wnt signalling in mice. *Nature* **407**, 535–538 (2000).
58. Mao, J. *et al.* Low-density lipoprotein receptor-related protein-5 binds to Axin and regulates the canonical Wnt signaling pathway. *Mol. Cell* **7**, 801–809 (2001).
59. Bain, G., Muller, T., Wang, X. & Papkoff, J. Activated β -catenin induces osteoblast differentiation of C3H10T1/2 cells and participates in BMP2 mediated signal transduction. *Biochem. Biophys. Res. Commun.* **301**, 84–91 (2003).
60. Zorn, A. M. Wnt signalling: antagonistic Dickkops. *Curr. Biol.* **11**, R592–R595 (2001).
61. Hadjiargyrou, M. *et al.* Transcriptional profiling of bone regeneration. Insight into the molecular complexity of wound repair. *J. Biol. Chem.* **277**, 30177–30182 (2002).
62. Van Wesenbeeck, L. *et al.* Six novel missense mutations in the LDL receptor-related protein 5 (*LRP5*) gene in different conditions with an increased bone density. *Am. J. Hum. Genet.* **72**, 763–771 (2003).
63. Balemans, W. *et al.* Increased bone density in sclerosteosis is due to the deficiency of a novel secreted protein (SOST). *Hum. Mol. Genet.* **10**, 537–543 (2001).
64. Brunkow, M. E. *et al.* Bone dysplasia sclerosteosis results from loss of the *SOST* gene product, a novel cystine knot-containing protein. *Am. J. Hum. Genet.* **68**, 577–589 (2001).
65. Balemans, W. & Van Hul, W. Extracellular regulation of BMP signaling in vertebrates: a cocktail of modulators. *Dev. Biol.* **250**, 231–250 (2002).
66. Rodan, G. A. & Martin, T. J. Therapeutic approaches to bone diseases. *Science* **289**, 1508–1514 (2000).
67. Mao, B. *et al.* Kremen proteins are Dickkopf receptors that regulate Wnt/ β -catenin signalling. *Nature* **417**, 664–667 (2002).

Acknowledgements We thank C. Hauer and J. Campbell for their assistance with artwork.

Evolving concepts of rheumatoid arthritis

Gary S. Firestein

Division of Rheumatology, Allergy and Immunology, School of Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-0656, USA (e-mail: gfirestein@ucsd.edu)

Rheumatoid arthritis is the most common inflammatory arthritis and is a major cause of disability. It existed in early Native American populations several thousand years ago but might not have appeared in Europe until the 17th century. Early theories on the pathogenesis of rheumatoid arthritis focused on autoantibodies and immune complexes. T-cell-mediated antigen-specific responses, T-cell-independent cytokine networks, and aggressive tumour-like behaviour of rheumatoid synovium have also been implicated. More recently, the contribution of autoantibodies has returned to the forefront. Based on the pathogenic mechanisms, specific therapeutic interventions can be designed to suppress synovial inflammation and joint destruction in rheumatoid arthritis.

"Life can only be lived forwards, but must be understood backwards." Søren Kierkegaard

Rheumatoid arthritis (RA) is a symmetric polyarticular arthritis that primarily affects the small diarthrodial joints of the hands and feet. In addition to inflammation in the synovium, which is the joint lining, the aggressive front of tissue called pannus invades and destroys local articular structures. The synovium is normally a relatively acellular structure with a delicate intimal lining. In RA, CD4⁺ T cells, B cells and macrophages infiltrate the synovium and sometimes organize into discrete lymphoid aggregates with germinal centres (Fig. 1). Hyperplasia of the intimal lining results from a marked increase in macrophage-like and fibroblast-like synoviocytes. Locally expressed degradative enzymes, including metalloproteinases, serine proteases and aggrecanases, digest the extracellular matrix and destroy the articular structures. RA occurs in 0.5–1.0% of the adult population worldwide, although the prevalence may have changed substantially in Europe since the Renaissance (see Box 1).

Era of rheumatoid factor and immune complexes

Prevailing notions on the pathogenesis of RA have evolved since the mid-20th century, when the first concepts of immune hyper-reactivity were considered. The first clue that self-reactivity plays a key role in RA was the identification of 'rheumatoid factor' in the blood of affected patients. Rheumatoid factor was observed originally by Waaler in 1939 and later rediscovered by Rose in 1948 by virtue of its ability to agglutinate sheep red cells that had been coated with rabbit serum. The seminal studies of Kunkel ultimately characterized the unknown factor as an antibody that binds to the Fc portion of immunoglobulins¹. This observation led to the logical view that RA might be an autoimmune disease caused by self-reactive antibodies. About 80% of patients are 'seropositive' for rheumatoid factor, and its presence predicts a more aggressive, destructive course.

The primary pathogenic potential of rheumatoid factor in RA as an initiator of immune complex-mediated disease was formulated during the 1960s and classically described by Zvaifler in 1973². In this model, immune complexes formed by rheumatoid factors and perhaps other autoantibodies fix

complement and release chemotactic factors such as C5a. Inflammatory cells are subsequently recruited to the rheumatoid joint along a chemotactic gradient where they are activated and contribute to local destruction. Neutrophils, in particular, accumulate in synovial fluid where they engulf immune complexes and release proteolytic enzymes.

Considerable data have accumulated to support this hypothesis. Ultrastructural studies of cartilage reveal immune complexes embedded in the superficial layers, thereby providing a solid surface to facilitate neutrophil adherence and activation. The synovium itself is a rich source for the local production of complement proteins and rheumatoid factor³. Despite the abundant production of complement in RA, synovial fluid concentrations are low in RA compared with other inflammatory arthropathies owing to intra-articular complement consumption⁴. However, many normal individuals and patients with other chronic inflammatory diseases produce rheumatoid factor, indicating that the mere presence of the autoantibody is insufficient for a phlogistic response. Because many immune-mediated and infectious diseases are marked by the presence of immune complexes, this hypothesis did not necessarily provide an adequate explanation of the unique features of RA.

Role of T cells in rheumatoid arthritis

Although the immune-complex theory could explain many of the acute inflammatory features of RA, the prominent T-cell infiltrate suggested these cells are key participants. In 1976, Stastny demonstrated that RA lymphocytes proliferate normally in allogeneic mixed-leukocyte reactions when stimulated by normal lymphocytes⁵, but the patients had deficient responses when stimulated by cells from other RA patients. These experiments actually demonstrated genetic similarities between RA patients and led to the discovery that specific human leukocyte antigen (HLA)-DR genes, which reside in the major histocompatibility complex (MHC) and participate in antigen presentation, are associated with the disease. The relationship between the MHC and RA is by far the strongest genetic link in RA and has been mapped precisely to the third hypervariable region of DRβ chains,

Box 1

Origins of rheumatoid arthritis

Examination of skeletal remains from antiquity in Europe and North Africa shows various forms of arthritis, including osteoarthritis, ankylosing spondylitis and gout⁶. But characteristic rheumatoid lesions with marginal erosions at the bone–cartilage interface of the small joints are strikingly absent. In contrast, palaeopathological studies of specimens dating back several thousand years show clear evidence of rheumatoid arthritis (RA) in Native American tribes in North America⁷⁷. The prevalence of RA in the same regions today remains extraordinarily high, with over 5% of individuals affected in some groups.

Evidence of RA in Europe first appeared in early 17th century art, especially by the Dutch Masters (see Box 1 Figure below), and Sydenham published the first case report in 1676. Although intermittent case series were subsequently reported, the disease was not fully recognized until it was defined by Garrod in 1859. He named it 'rheumatoid' arthritis to distinguish it from the two well-known forms arthritis, rheumatic fever and gout. By the early 20th century, RA was viewed as separate from osteoarthritis ('arthritis deformans'). In 1957, Charles Short described RA definitively and clearly set it apart as a defined clinical entity distinct from the seronegative spondyloarthropathies, crystal-induced disease, osteoarthritis, systemic lupus erythematosus, and many other conditions.



by Jacob Jordaens (c. 1630) is shown. Note swelling of the metacarpal-phalangeal and proximal interphalangeal joints (arrows). (Image courtesy of the Museo del Prado, Madrid.)

Box 1 Figure Example of RA-like findings in Dutch art. Little evidence of RA-like disease was noted in art or skeletal remains before the 17th century in Europe and Northern Africa. Findings suggestive of RA appear in 17th century Dutch art. Detail from *La Familia de Jordaens en un Jardín*

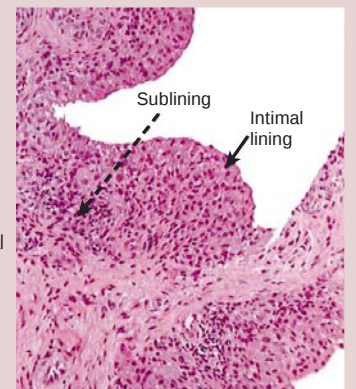


especially amino acids 70 through 74 (ref. 6). The susceptibility epitope is glutamine-leucine-arginine-alanine-alanine (QKRAA) or QRRAA. This sequence is found in multiple RA-associated DR genes, including DR4, DR14 and DR1 (for example, DRB*0401, DRB*0404, DRB*0101 and DRB*1402). The susceptibility epitope might also influence severity of disease, as the risk of extra-articular and erosive disease is greater if patients have the gene and is further increased by homozygosity⁷.

The location of the susceptibility epitope on the MHC molecule suggests that it might have a role in the ability of HLA-DR to bind and present specific arthritogenic peptides that might cause RA. The precise role of HLA-DR in RA could be more complex, as crystal structures of DR molecules show that the susceptibility epitope residues primarily face away from the antigen-binding groove of DR molecules that determine which peptides are presented to CD4⁺ T cells⁸. Specific peptides that bind to the DR proteins in RA patients have not been identified⁹. Several alternative roles for the role of the susceptibility epitope have been proposed (see Table 1).

The notion that T cells participate in the aetiology and pathogenesis of RA was a key intellectual breakthrough in our understanding of the disease¹⁰. Several animal models of arthritis commonly used to investi-

Figure 1 Synovial histology in rheumatoid arthritis. A photomicrograph (magnification $\times 200$) shows the redundant folds of the synovial lining and intense infiltration with inflammatory cells in RA. The intimal lining layer (solid arrow) is hyperplastic, with multiple layers of cells compared with a normal lining that is one or two cell layers deep. The sublining region (dashed arrow) is marked by accumulation of mononuclear cells such as CD4⁺ T cells, macrophages and B cells.



gate the mechanisms of synovitis, such as collagen-induced arthritis or adjuvant arthritis in rodents, are clearly T-cell-dependent and provide a link between pre-clinical and clinical investigation. This evidence is circumstantial, however, and the arthritogenicity of an antigen in mice is no guarantee that it has a role in the human disease. Using *in vitro* proliferation assays with T cells from RA patients, many putative autoantigens have been identified as candidate antigens. Type II collagen, proteoglycans, aggrecan, cartilage link protein, heat shock proteins and many relatively specific joint antigens have been implicated by virtue of the T-cell reactivity^{11,12} and antibodies^{13,14} in RA, as well as their ability to cause arthritis in appropriately immunized mice¹⁵.

Explaining disease perpetuation using cytokine networks

Cytokine autocrine and paracrine networks

By the late 1980s, new molecular techniques were available to measure cytokines in RA synovium and synovial fluid, which offered an opportunity to confirm the role of T cells as the driving force in RA. The most striking observation was that T-cell cytokines such as interleukin (IL)-2 and interferon (IFN)- γ were present in relatively low concentrations, whereas macrophage and fibroblast products were abundant¹⁶. Subsequent studies confirmed the unbalanced cytokine profile and indicated that T-cell lymphokines, although detectable, were present in low amounts compared with other antigen-mediated processes like tuberculous pleuritis, asthma or inflamed tonsils¹⁷. In contrast, a broad array of macrophage and fibroblast cytokines, including IL-1, IL-6, IL-15, IL-18, tumour-necrosis factor (TNF)- α , granulocyte-macrophage colony-stimulating factor (GM-CSF), various chemokines, and many others, are produced by rheumatoid synovium. Interrupting cytokine networks to treat RA using biological agents, such as anti-TNF- α antibody, was supported by subsequent studies demonstrating efficacy in collagen-induced arthritis in mice¹⁸. Suppressive cytokines, such as transforming growth factor (TGF)- β and IL-1Ra, as well as anti-inflammatory cytokine signalling mechanisms, such as the suppressor of cytokine signalling-3 (SOCS3), are expressed in RA synovium, but at levels that are inadequate to block synovitis¹⁹.

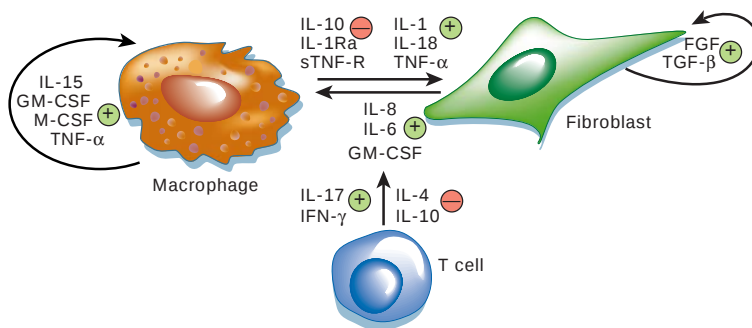
To explain this apparent paradox, paracrine and autocrine networks involving factors such as IL-1 and TNF- α were proposed in 1990 as mechanisms that perpetuate synovitis in a T-cell-in

Table 1 Potential role of HLA-DR in RA^{72–75}

• Binds to arthritogenic peptides that can be presented to T cells
• Shapes the T-cell repertoire and permits escape from tolerance or survival of autoreactive clones
• Leads to enhanced T-cell reactivity owing to unique contacts between T-cell receptors and MHC molecules
• Serves as a target for autoreactive T cells owing to molecular mimicry with a pathogen (for example, <i>Escherichia coli</i> DnaJ or Epstein–Barr virus peptides)
• Closely linked to other genes in the MHC that are associated with RA
• Fails to bind peptides of an arthritogenic pathogen, leading to an inadequate immune response

Figure 2 Cytokine networks in rheumatoid arthritis.

Macrophages and fibroblasts are adjacent to one another in the synovial intimal lining and produce cytokines that can activate either themselves or their neighbouring cells. Pro-inflammatory cytokines (+) and anti-inflammatory proteins (–) are indicated. This scenario involving interactions between mesenchymal cells and antigen-presenting cells can participate in many inflammatory reactions and is not necessarily specific to RA. The process is not autonomous because cessation of anti-cytokine therapy leads to flares of disease in RA.



dependent manner²⁰. Fibroblast-like synoviocytes and macrophage-like synoviocytes in the intimal lining can produce factors that maintain inflammation by activating cells in the local environs (Fig. 2). Adjacent cells express other cytokines, metalloproteinases, and small molecule mediators such as prostaglandins and nitric oxide. Adhesion molecules are induced on endothelium, and circulating cells expressing the appropriate chemokine receptors and adhesion molecule receptors accumulate in the synovium. This model could explain the accumulation of memory T_H1 cells, which express the chemokine receptor CCR5 and the integrin adhesion molecule $\alpha_4\beta_1$, and are the key effector cells in delayed-type hypersensitivity, as well as monocytes and neutrophils without necessarily invoking a specific antigen. The cells accumulate in the joint and might be activated by articular antigens that are unrelated to the aetiological factors.

Recent therapeutic interventions, including TNF- α and IL-1 inhibitors, strongly support the importance of cytokines in RA. The mechanism of anti-TNF- α action *in vivo* is complex and probably includes suppression of other pro-inflammatory cytokines, decreased synovial cellular infiltration, interference with osteoclast activation, and decreased angiogenesis²¹. Continuous anti-cytokine treatment is required for long-term control and the disease flares when therapy is discontinued. Therefore, a relatively simple view of autonomous cytokine networks cannot explain perpetuation of RA unless additional stimuli that maintain cytokine production are present. The role of TNF- α is also not unique to RA; TNF inhibitors have shown efficacy in Crohn's disease, ankylosing spondylitis, psoriasis and psoriatic arthritis. New factors have been implicated and are also potential therapeutic targets for RA. For instance, IL-18, which is a member of the IL-1 family and can be inhibited by IL-18-binding protein, activates synovial macrophages and biases T cells to differentiate towards the T_H1 phenotype²². IL-15 is produced by macrophages and can activate T cells to increase TNF- α production by macrophages.

More precise characterization of T-cell products and phenotype requires modification of the cytokine network hypothesis. Careful analysis of infiltrating synovial T lymphocytes has revealed a bias towards T_H1 -like cells that express CCR5 and CXCR3 chemokine receptors²³. Small but physiologically relevant amounts of IFN- γ and the T_H1 cytokine IL-17 are expressed in RA and could contribute to immune responses, fibroblast activation and bone destruction²⁴. Direct cell–cell contact between memory or previously activated T cells and other resident synovial cells can also enhance cytokine and metalloproteinase production²⁵. Therefore, chronic antigen-dependent and -independent T-cell activation could operate within the rheumatoid synovium.

Role of cytokines in bone destruction

Studies using anti-TNF agents clearly show that they can slow or prevent the progression of bone and cartilage damage in RA²⁶. This activity probably involves suppression of osteoclasts in joint lesions²⁷. Other cytokines regulating matrix degradation have also been impli-

cated in animal models of arthritis, most notably IL-1 and IL-17²⁸. Perhaps the most exciting development in the pathogenesis of bone erosions in RA relates to the discovery of osteoclast-mediated bone resorption that is regulated by RANKL, the RANK (receptor activator of nuclear factor (NF)- κ B) ligand. RANKL is expressed by a variety of cell types involved in RA, including T cells and synoviocytes. These cells, in the presence of cytokines like TNF- α and M-CSF, contribute to osteoclast maturation and activation. The soluble decoy receptor to RANKL, osteoprotegerin (OPG), and RANKL are increased in RA, but normalize after treatment with TNF inhibitors²⁹.

The role of RANKL in inflammatory joint disease has been confirmed in several animal models. For instance, T-cell activation leads to a RANKL-mediated increase in osteoclasts and bone loss in rat adjuvant arthritis³⁰. OPG administration to the arthritic animals blocks bone destruction even though it has little effect on inflammation. RANKL-knockout mice also have diminished bone erosion in arthritis models³¹. In most cases, cartilage protection afforded by blocking RANK/RANKL system is minimal.

Mechanisms of cytokine gene expression and new therapeutic targets

Understanding the intracellular targets that regulate cytokines in RA can potentially lead to new therapeutic interventions. For instance, NF- κ B is activated in the synovium of patients with RA³² and regulates genes that contribute to inflammation, including TNF- α , IL-6, IL-8, inducible nitric oxidase synthase (iNOS) and cyclooxygenase-2 (COX-2). After stimulation of innate immunity or exposure to pro-inflammatory cytokines, the I κ B kinase (IKK) signal complex is activated in synoviocytes, leading to phosphorylation of I κ B³³. IKK β is both necessary and sufficient for induction of IL-6, IL-8 and intercellular adhesion molecule-1 (ICAM-1) gene expression. Targeting NF- κ B is an effective therapeutic strategy in many animal models of arthritis. For instance, rat adjuvant-induced arthritis is suppressed by intra-articular gene therapy with dominant negative IKK β adenoviral construct³⁴, while decoy oligonucleotides block streptococcal cell-wall arthritis³⁵.

The mitogen-activated protein (MAP) kinases are also key regulators of cytokine and metalloproteinase production and could also be targeted in RA. All three kinase families — extracellular signal-regulated kinase (ERK), c-Jun N-terminal kinase (JNK) and p38 — are expressed in rheumatoid synovial tissue³⁶. All are constitutively expressed by cultured synoviocytes, and exposure to pro-inflammatory cytokines induces rapid phosphorylation. Upstream kinases that activate the MAP kinases, such as MKK3, MKK4, MKK6 and MKK7, are also activated in RA synovium and can form signalling complexes that integrate external environmental stresses to generate an appropriate cellular response.

The MAP kinases have attracted considerable attention as potential therapeutic targets in RA. Pre-clinical studies show that p38 inhibitors are effective in murine collagen-induced arthritis, rat adjuvant-induced arthritis, and many other models³⁷. A selective JNK inhibitor, SP600125, is mildly anti-inflammatory in the rat

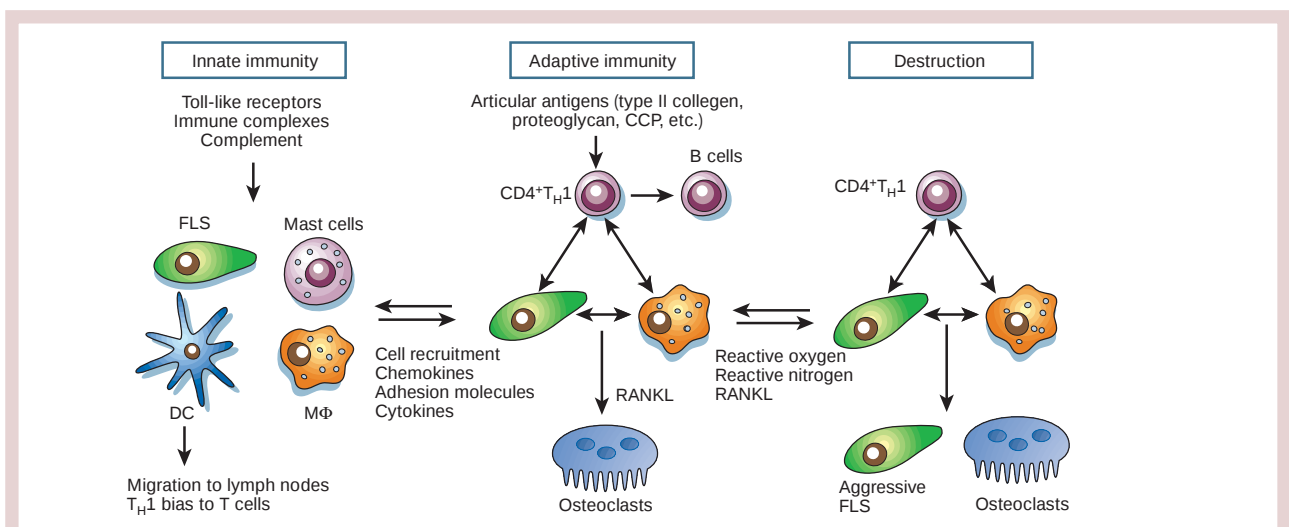


Figure 3 A proposed model implicating multiple pathogenic mechanisms in RA.

According to this model, a step-wise progression can begin with the activation of innate immunity by stimulating dendritic cells, macrophages, fibroblasts and mast cells. After immune cells migrate into the synovium, an opportunity for adaptive immune responses arises in individuals with the appropriate genetic background. While antigen presentation can occur in the synovium, extra-articular sites can also participate if dendritic cells migrate to lymph nodes and bias T cells to a T_H1 phenotype. In the

destructive phase of disease, osteoclast activation mediates abundant bone resorption under the influence of RANKL, while synoviocytes can invade cartilage. These processes are not necessarily mutually exclusive. Activation of innate and adaptive immunity can also occur in a parallel fashion (denoted by bidirectional arrows), perhaps contributing to the patterns of disease flares and remissions. DC, dendritic cell; CCP, cyclic citrullinated peptide; FLS, fibroblast-like synoviocyte; MΦ, macrophage.

adjuvant-induced arthritis, but provides striking protection against bone and cartilage destruction³⁸. Of the JNK isoforms, JNK2 is particularly important in arthritis because it is the dominant isoform expressed in synoviocytes. In JNK2-knockout mice, passive collagen-induced arthritis causes less cartilage damage compared with wild-type animals although clinical arthritis is still severe³⁹.

The transcription factor activator protein-1 (AP-1) also regulates many genes that participate in RA, including TNF- α and metalloproteinases. High levels of AP-1 binding activity are detected in nuclear extracts of RA synovial tissue compared with osteoarthritis. Its components c-Jun and c-Fos are highly expressed in RA synovium, especially in the nuclei of cells in the intimal lining layer⁴⁰. Pro-inflammatory cytokines can activate AP-1 activity in synoviocytes and leads to massive release of metalloproteinases. AP-1 decoy oligonucleotides suppress collagen-induced arthritis and inhibit IL-1, IL-6, TNF- α , matrix metalloproteinase (MMP)-3 and MMP-9 production by synovial tissue⁴¹.

One concern with therapy directed at any of these key regulatory pathways is that they also participate in many normal cellular functions. The risks of toxicity or impaired host defence are significant potential problems because alterations in innate immunity or adaptive responses can be suppressed.

Rheumatoid arthritis as a locally invasive tumour

Autoimmune and inflammatory mechanisms account for many features of RA. However, relentless progression of joint destruction in the relative absence of synovitis contributed to a parallel line of reasoning in the 1990s that viewed aggressive characteristics of RA synovium as reminiscent of neoplastic tissue. For instance, RA synoviocytes can grow under anchorage-independent conditions and have defective contact inhibition. Examination of X-linked genes demonstrates oligoclonality in the synoviocyte population from RA, but not osteoarthritis, synovium⁴². Permanent changes in synoviocyte function are also apparent in the severe combined immunodeficiency (SCID) mouse model where RA, but not osteoarthritis, synoviocytes invade into cartilage explants⁴³. Synovial dedifferentiation has also been described, including overexpression of embryonic genes such as *wnt5A* that can enhance IL-15 expression

in rheumatoid synoviocytes⁴⁴. As in neoplastic disease, angiogenesis feeds the expanding synovium⁴⁵.

The molecular mechanisms responsible for altered synoviocyte function in RA could include somatic mutations in key genes. The presence of mutations in genes like the p53 tumour suppressor, although controversial, has been identified in RA synovial tissue and synoviocytes^{46–48}. Microsatellite instability, another indicator of DNA damage, also occurs in RA synovial tissue, and may be related to decreased expression of DNA mismatch repair enzymes such as hMSH6⁴⁹. Some of the p53 mutations identified in RA are dominant negative and are unable to regulate IL-6 or *bax* gene expression. Microdissection of RA synovium localizes islands of p53 mutant cells to the intimal lining, with higher expression of IL-6 than wild-type regions⁵⁰. DNA damage in RA and resultant mutations in regulatory genes likely results from exposure to persistent oxidative stress in a hostile environment. The mutations, therefore, are the result of RA rather than the cause and potentially alter the natural history of RA. However, the ultimate impact of relatively small numbers of abnormal cells on disease progression is uncertain.

Return of immune complexes and autoantibodies

The recent fortuitous discovery of a spontaneous arthritis model in mice that produce antibodies directed against a ubiquitous antigen, glucose-6-phosphoisomerase (GPI), contributed to resurgent interest in autoantibodies and immune complexes⁵¹. The murine arthritis can be transferred using serum from affected mice. Elegant molecular studies using knockout mice demonstrate an absolute requirement for Fc γ R (especially Fc γ RIII) and components of the alternative complement cascade (for example, Factor B) as well as complement proteins C3 and C5⁵². Although the model is, in part, dependent on TNF- α , IL-1 plays a pivotal role.

Although anti-GPI antibodies are present in many patients, they are neither specific nor sensitive for RA⁵³. The ability of ubiquitous antigens like GPI to induce synovial inflammation is probably related to their adherence to the cartilage surfaces. The presentation of immobilized antigen-antibody complexes on cartilage provides an exceptionally good substrate for complement fixation, similar to rheumatoid factor embedded in rheumatoid cartilage. A similar situation occurs in

murine passive collagen-induced arthritis, where anti-type II collagen antibody binds to collagen arrayed on the surface of articular cartilage.

Many autoantibody systems that could participate in inflammatory joint disease are now recognized in RA, including rheumatoid factor, type II collagen, immunoglobulin heavy chain binding protein (BiP), heat shock proteins and hnRNP-33 (RA33)⁵⁴. Some, like anti-collagen antibody and rheumatoid factor, are produced by rheumatoid synovial B cells. The mere presence of the autoantibodies in serum is not specific for RA; instead, patterns of antigen immunoreactivity could be pathogenic in humans (a form of autoantibody 'signature'). Antibodies directed against proteins containing citrulline, which is derived from post-translational modification of arginine by peptidylarginine deiminase, might be more specific for RA⁵⁵.

Careful analysis of the B-cell repertoire in rheumatoid synovial tissue suggests a local antigen-driven immune response. Most B cells isolated from germinal centres in RA synovium have unmutated *VH* genes, suggesting that they have recently migrated into the tissue and are activated at the site of disease⁵⁶. Shared mutations containing identical sequences throughout the variable domain occur in RA synovial B cells, and a limited number of *VH* and *DH* gene segments are utilized⁵⁷. The subsequent gene rearrangements and mutations implicate synovial autoimmunity directed against articular antigens.

Innate immunity also participates in antibody-mediated arthritis, and mast cells are required for the disease onset of some models⁵⁸. Toll-like receptors (TLRs) can also participate. For instance, administration of a small dose of endotoxin, which activates TLR4, during the pre-clinical phase of collagen-induced arthritis can markedly exacerbate joint inflammation and destruction⁵⁹. CpG oligonucleotides can engage TLR9 and either cause arthritis directly in rodents or enhance disease in adjuvant-induced arthritis⁶⁰.

From bench to bedside

The advent of TNF inhibitors^{61,62} illustrates the success of applied translational research in RA, based on characterization of cytokine networks and studies suggesting that TNF- α production might serve as an autologous stimulus for other cytokines in RA synovium⁶³. Although perhaps 40% of patients have dramatic responses, the remainder have some evidence of persistent synovitis or minimal clinical benefit. IL-1Ra, a natural IL-1 antagonist, has also been approved for use in the United States. The response rates are less than for TNF inhibitors, perhaps because IL-1Ra is a competitive antagonist that must be present in large excess. Numerous additional cytokine-directed agents, such as anti-IL-6 receptor antibody, are also in clinical development, with preliminary response rates similar to TNF antagonists. Inhibition of IL-18 and IL-15 represents additional attractive approaches that could block T_H1 differentiation, inflammatory mediator production, or TNF- α expression. Based on the T_H1 bias of T cells in the synovium, treatment with T_H2 cytokines (IL-4, IL-10 and IL-13) was tested in many animal models of arthritis and showed considerable promise⁶⁴. But so far, administration of IL-10, which serves as a prototypic T_H2 cytokine, has met with only limited success⁶⁵.

Our understanding of the signal transduction pathways implicated in RA has led to drug development programmes targeting MAP kinase and NF- κ B inhibitors. In addition, biological agents like CTLA4-Ig, designed to interfere with co-stimulatory molecules and antigen presentation, have been impressive and raise the possibility that modulation of T-cell function will have therapeutic utility in RA. Adhesion molecule inhibitors (for example, $\alpha_4\beta_1$ integrin) or chemokine receptor antagonists (for example, CCR1 or CCR5) could block cell ingress into affected synovial tissue. OPG might prevent or heal bone erosions by blocking the differentiation and activation of osteoclasts. The extracellular matrix could also be protected by inhibiting metalloproteinases (for example, MMP13 or collagenase-3), aggrecanases or cathepsins (for example, cathepsin K). Clinical studies with broad-spectrum metalloproteinase inhibitors have demonstrated limited effect on joint destruction and have been complicated by some toxicity. The tumour-like properties of synovium

suggest that approaches normally reserved for oncologic diseases might be useful in RA, including induction of apoptosis (for example, Fas ligation)⁶⁶ and anti-angiogenesis agents.

Finally, the role of autoantibodies raises the possibility that B-cell-specific therapy might be useful in RA. Rituximab, which targets CD20, seems to be effective in RA, especially in combination with cyclophosphamide, methotrexate and/or corticosteroids⁶⁷. The mechanism of action is still uncertain and might involve inhibition of B-cell cytokines or antigen presentation, because changes in the amount of rheumatoid factor and autoantibody do not necessarily correlate with clinical responses. Other cytokines that are important in B-cell differentiation, like BlyS, are also potential targets⁶⁸. The use of soluble receptors to BlyS and related proteins is effective in collagen-induced arthritis. Inhibition of autoantibody production through these mechanisms or blockade of innate immunity, such as with C5a antagonists, have also attracted interest⁶⁹.

Updated model incorporating 50 years of theories

The various hypotheses that have attempted to explain RA over the last half century might be re-interpreted in order to appreciate the complexity of the disease and the multiple mechanisms of synovial inflammation. A multi-stage model might help integrate these concepts, although it is important to note that significant temporal overlap can occur (see Fig. 3)⁷⁰. Activation of innate immunity probably occurs early in RA and can serve as a key pathogenic mechanism that initiates synovial inflammation. TLR agonists, including proteoglycans and bacterial DNA, have been detected in rheumatoid synovium⁷¹. Engagement of Fc receptors by autoantibodies, like rheumatoid factor and anti-cyclic citrullinated peptide antibodies, represents an alternative mechanism. Synovial dendritic cells activated by TLR ligands can migrate to lymph nodes where primed T cells can be biased towards the T_H1 phenotype and, through chemokine receptors like CCR5, home to inflamed synovial tissue. Production of cytokines and expression of adhesion molecules after activation of innate immunity in the joint can permit the continued ingress of immune cells. This early stage might occur commonly in humans but is typically self-limited.

In the appropriate conditions, such as the presence of the correct HLA-DR alleles or perhaps cytokine promoter polymorphisms, lymphocytes could accumulate into the transiently inflamed synovium. Under these circumstances, the loss of tolerance related to the effects of HLA-DR background or the T-cell repertoire might contribute to autoreactivity to newly exposed articular antigens. Ultimately, long-standing disease could lead to a destructive phase that some investigators believe can run an independent course.

This hypothesis does not presuppose a specific 'rheumatoid antigen'. Instead, a variety of antigens can be targets and lead to both T-cell activation and B-cell maturation. Hence, the progression of disease might require a combination of stochastic events (activation of innate immunity), pre-determined events (genetic background), and adaptive immune responses directed against autologous antigens. It is also possible that recurrent activation of innate immunity or overlap with other elements of this model could participate to varying degrees during the course of disease and contribute to patterns of flare and remission.

Although significant therapeutic advances have greatly improved the lives of patients with RA, the problem is hardly solved. Persistent disease and the accumulated disability caused by decades of inflammatory synovitis remains a major clinical problem. By targeting the intracellular signalling pathways, cytokines, and other mediators of the phases of RA, the residual disease could be controlled. Perhaps with appropriate therapy the impact of RA will fade to the halcyon days of pre-Renaissance Europe when the disease was essentially unknown. □

doi:10.1038/nature01661

- Franklin, E. C., Molman, H. R., Muller-Eberhard, H. J. & Kunkel, H. B. An unusual protein component of high molecular weight in the serum of certain patients with rheumatoid arthritis. *J. Exp. Med.* **105**, 425–435 (1957).

2. Zvaifler, N. J. The immunopathology of joint inflammation in rheumatoid arthritis. *Adv. Immunol.* **16**, 265–336 (1973).
3. Neumann, E. *et al.* Local production of complement proteins in rheumatoid arthritis synovium. *Arthritis Rheum.* **46**, 934–945 (2002).
4. Ruddy, S. & Austen, K. F. Activation of the complement and properdin systems in rheumatoid arthritis. *Ann. NY Acad. Sci.* **256**, 96–104 (1975).
5. Stastny, P. Mixed lymphocyte cultures in rheumatoid arthritis. *J. Clin. Invest.* **57**, 1148–1157 (1976).
6. Nepom, G. T. *et al.* HLA genes associated with rheumatoid arthritis: identification of susceptibility alleles using specific oligonucleotide probes. *Arthritis Rheum.* **32**, 15–21 (1989).
7. Weyand, C. M., Hicok, K. C., Conn, D. L. & Goronzy, J. J. The influence of HLA-DRB1 genes on disease severity in rheumatoid arthritis. *Ann. Intern. Med.* **117**, 801–803 (1992).
8. Stern, L. J. *et al.* Structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. *Nature* **368**, 215–221 (1994).
9. Kirschmann, D. A. *et al.* Naturally processed peptides from rheumatoid arthritis and non-associated HLA-DR alleles. *J. Immunol.* **155**, 5655–5662 (1995).
10. Fox, D. A. The role of T cells in the immunopathogenesis of rheumatoid arthritis: new perspectives. *Arthritis Rheum.* **40**, 598–609 (1997).
11. Verheijden, G. F. *et al.* Human cartilage glycoprotein-39 as a candidate autoantigen in rheumatoid arthritis. *Arthritis Rheum.* **40**, 1115–1125 (1997).
12. Li, N. L. *et al.* Isolation and characteristics of autoreactive T cells specific to aggrecan G1 domain from rheumatoid arthritis patients. *Cell Res.* **10**, 39–49 (2000).
13. Oda, A. *et al.* Antibodies to 65Kd heat-shock protein were elevated in rheumatoid arthritis. *Clin. Rheumatol.* **13**, 261–264 (1994).
14. Rowley, M., Tai, B., Mackay, I. R., Cunningham, T. & Phillips, B. Collagen antibodies in rheumatoid arthritis. Significance of antibodies to denatured collagen and their association with HLA-DR4. *Arthritis Rheum.* **29**, 174–184 (1986).
15. Courtenay, J. S., Dallman, M. J., Dayan, A. D., Martin, A. & Mosedale, B. Immunisation against heterologous type II collagen induces arthritis in mice. *Nature* **283**, 666–668 (1980).
16. Firestein, G. S., Alvaro-Gracia, J. M. & Maki, R. Quantitative analysis of cytokine gene expression in rheumatoid arthritis. *J. Immunol.* **144**, 3347–3353 (1990).
17. Smeets, T. J. *et al.* Poor expression of T cell-derived cytokines and activation and proliferation markers in early rheumatoid synovial tissue. *Clin. Immunol. Immunopathol.* **88**, 84–90 (1998).
18. Williams, R. O., Mason, L. J., Feldmann, M. & Maini, R. N. Synergy between anti-CD4 and anti-tumor necrosis factor in the amelioration of established collagen-induced arthritis. *Proc. Natl Acad. Sci. USA* **91**, 2762–2766 (1994).
19. Shouda, T. *et al.* Induction of the cytokine signal regulator SOCS3/CIS3 as a therapeutic strategy for treating inflammatory arthritis. *J. Clin. Invest.* **108**, 1781–1788 (2001).
20. Firestein, G. S. & Zvaifler, N. J. How important are T cells in chronic rheumatoid synovitis? *Arthritis Rheum.* **33**, 768–773 (1990).
21. Feldmann, M. & Maini, R. N. Anti-TNF α therapy of rheumatoid arthritis: what have we learned? *Annu. Rev. Immunol.* **19**, 163–196 (2001).
22. Gracie, J. A. *et al.* A proinflammatory role for IL-18 in rheumatoid arthritis. *J. Clin. Invest.* **104**, 1393–1401 (1999).
23. Qin, S. *et al.* The chemokine receptors CXCR3 and CCR5 mark subsets of T cells associated with certain inflammatory reactions. *J. Clin. Invest.* **101**, 746–754 (1998).
24. Chabaud, M., Fossiez, F., Taupin, J. L. & Miossec, P. Enhancing effect of IL-17 on IL-1-induced IL-6 and leukemia inhibitory factor production by rheumatoid arthritis synoviocytes and its regulation by Th2 cytokines. *J. Immunol.* **161**, 409–414 (1998).
25. Burger, D. *et al.* Imbalance between interstitial collagenase and tissue inhibitor of metalloproteinases 1 in synoviocytes and fibroblasts upon direct contact with stimulated T lymphocytes: involvement of membrane-associated cytokines. *Arthritis Rheum.* **41**, 1748–1759 (1998).
26. Lipsky, P. E. *et al.* Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. *N. Engl. J. Med.* **343**, 1594–1602 (2000).
27. Redlich, K. *et al.* Osteoclasts are essential for TNF- α -mediated joint destruction. *J. Clin. Invest.* **110**, 1419–1427 (2002).
28. Lubberts, E. *et al.* IL-17 promotes bone erosion in murine collagen-induced arthritis through loss of the receptor activator of NF- κ B ligand/osteoprotegerin balance. *J. Immunol.* **170**, 2655–2662 (2003).
29. Ziolkowska, M. *et al.* High levels of osteoprotegerin and soluble receptor activator of nuclear factor κ B ligand in serum of rheumatoid arthritis patients and their normalization after anti-tumor necrosis factor α treatment. *Arthritis Rheum.* **46**, 1744–1753 (2002).
30. Kong, Y. Y. *et al.* Activated T cells regulate bone loss and joint destruction in adjuvant arthritis through osteoprotegerin ligand. *Nature* **402**, 304–309 (1999).
31. Pettit, A. R. *et al.* TRANCE/RANKL knockout mice are protected from bone erosion in a serum transfer model of arthritis. *Am. J. Pathol.* **159**, 1689–1699 (2001).
32. Han, Z., Boyle, D. L., Manning, A. M. & Firestein, G. S. AP-1 and NF- κ B regulation in rheumatoid arthritis and murine collagen-induced arthritis. *Autoimmunity* **28**, 197–208 (1998).
33. Aupperle, K. *et al.* NF- κ B regulation by I κ B kinase-2 in rheumatoid arthritis synoviocytes. *J. Immunol.* **166**, 2705–2711 (2001).
34. Tak, P. P. *et al.* Inhibitor of nuclear factor κ B kinase β is a key regulator of synovial inflammation. *Arthritis Rheum.* **44**, 1897–1907 (2001).
35. Miagkov, A. V. *et al.* NF- κ B activation provides the potential link between inflammation and hyperplasia in the arthritic joint. *Proc. Natl Acad. Sci. USA* **95**, 13859–13864 (1998).
36. Schett, G. *et al.* Activation, differential localization, and regulation of the stress-activated protein kinases, extracellular signal-regulated kinase, c-JUN N-terminal kinase, and p38 mitogen-activated protein kinase, in synovial tissue and cells in rheumatoid arthritis. *Arthritis Rheum.* **43**, 2501–2512 (2000).
37. Badger, A. M. *et al.* Pharmacological profile of SB 203580, a selective inhibitor of cytokine suppressive binding protein/p38 kinase, in animal models of arthritis, bone resorption, endotoxin shock and immune function. *J. Pharmacol. Exp. Ther.* **279**, 1453–1461 (1996).
38. Han, Z. *et al.* c-Jun N-terminal kinase is required for metalloproteinase expression and joint destruction in inflammatory arthritis. *J. Clin. Invest.* **108**, 73–81 (2001).
39. Han, Z., Chang, L., Yamanishi, Y., Karin, M. & Firestein, G. S. Joint damage and inflammation in c-Jun N-terminal kinase 2 knockout mice with passive murine collagen-induced arthritis. *Arthritis Rheum.* **46**, 818–823 (2002).
40. Kinne, R. W. *et al.* Synovial fibroblast-like cells strongly express jun-B and C-fos proto-oncogenes in rheumatoid- and osteoarthritis. *Scand. J. Rheumatol. Suppl.* **101**, 121–125 (1995).
41. Shiozawa, S., Shimizu, K., Tanaka, K. & Hino, K. Studies on the contribution of c-fos/AP-1 to arthritic joint destruction. *J. Clin. Invest.* **99**, 1210–1216 (1997).
42. Imamura, F. *et al.* Monoclonal expansion of synoviocytes in rheumatoid arthritis. *Arthritis Rheum.* **41**, 1979–1986 (1998).
43. Pap, T. *et al.* Invasiveness of synovial fibroblasts is regulated by p53 in the SCID mouse in vivo model of cartilage invasion. *Arthritis Rheum.* **44**, 676–681 (2001).
44. Sen, M., Chamorro, M., Reifert, J., Corr, M. & Carson, D. A. Blockade of Wnt-5A/Frizzled 5 signaling inhibits rheumatoid synovioyte activation. *Arthritis Rheum.* **44**, 772–781 (2001).
45. Walsh, D. A., Wade, M., Mapp, P. I. & Blake, D. R. Focally regulated endothelial proliferation and cell death in human synovium. *Am. J. Pathol.* **152**, 691–702 (1998).
46. Firestein, G. S. *et al.* Somatic mutations in the p53 tumor suppressor gene in rheumatoid arthritis synovium. *Proc. Natl Acad. Sci. USA* **94**, 10895–10900 (1997).
47. Reme, T. *et al.* Mutations of the p53 tumour suppressor gene in erosive rheumatoid synovial tissue. *Clin. Exp. Immunol.* **111**, 353–358 (1998).
48. Inazuka, M. *et al.* Analysis of p53 tumour suppressor gene somatic mutations in rheumatoid arthritis synovium. *Rheumatology* **39**, 262–266 (2000).
49. Lee, S.-H. *et al.* Microsatellite instability and suppressed DNA repair enzyme expression in rheumatoid arthritis. *J. Immunol.* **170**, 2214–2220 (2003).
50. Yamanishi, Y. *et al.* Regional analysis of p53 mutations in rheumatoid arthritis synovium. *Proc. Natl Acad. Sci. USA* **99**, 10025–10030 (2002).
51. Kouskoff, V. *et al.* Organ-specific disease provoked by systemic autoimmunity. *Cell* **87**, 811–822 (1996).
52. Ji, H. *et al.* Arthritis critically dependent on innate immune system players. *Immunity* **16**, 157–168 (2002).
53. Matsumoto, I. *et al.* Low prevalence of antibodies to glucose-6-phosphate isomerase in patients with rheumatoid arthritis and a spectrum of other chronic autoimmune disorders. *Arthritis Rheum.* **48**, 944–954 (2003).
54. Steiner, G. & Smolen, J. Autoantibodies in rheumatoid arthritis and their clinical significance. *Arthritis Res.* **4**, S1–S5 (2002).
55. Union, A. *et al.* Identification of citrullinated rheumatoid arthritis-specific epitopes in natural filaggrin relevant for antifilaggrin autoantibody detection by line immunoassay. *Arthritis Rheum.* **46**, 1185–1195 (2002).
56. Kim, H.-J. & Berek, C. B cells in rheumatoid arthritis. *Arthritis Res.* **2**, 126–131 (2000).
57. Clausen, B. E. *et al.* Clonally-related immunoglobulin VH domains and nonrandom use of DH gene segments in rheumatoid arthritis synovium. *Mol. Med.* **4**, 240–251 (1998).
58. Lee, D. M. *et al.* Mast cells: a cellular link between autoantibodies and inflammatory arthritis. *Science* **297**, 1689–1692 (2002).
59. Terato, K. *et al.* Collagen-induced arthritis in mice: synergistic effect of *E. coli* lipopolysaccharide bypasses epitope specificity in the induction of arthritis with monoclonal antibodies to type II collagen. *Autoimmunity* **22**, 137–147 (1995).
60. Ronaghy, A. *et al.* Immunostimulatory DNA sequences influence the course of adjuvant arthritis. *J. Immunol.* **168**, 51–56 (2002).
61. Moreland, L. W. *et al.* Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75)-Fc fusion protein. *N. Engl. J. Med.* **337**, 141–147 (1997).
62. Elliott, M. J. *et al.* Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor alpha (cA2) versus placebo in rheumatoid arthritis. *Lancet* **344**, 1105–1110 (1994).
63. Brennan, F. M., Chantray, D., Jackson, A., Maini, R. & Feldmann, M. Inhibitory effect of TNF α antibodies on synovial cell interleukin-1 production in rheumatoid arthritis. *Lancet* **2**, 244–247 (2001).
64. Lubberts, E. *et al.* IL-4 gene therapy for collagen arthritis suppresses synovial IL-17 and osteoprotegerin ligand and prevents bone erosion. *J. Clin. Invest.* **105**, 1697–1710 (2000).
65. Smeets, T. J., Kraan, M. C., Versendaal, J., Breedveld, F. C. & Tak, P. P. Analysis of serial synovial biopsies in patients with rheumatoid arthritis: description of a control group without clinical improvement after treatment with interleukin 10 or placebo. *J. Rheumatol.* **26**, 2089–2093 (1999).
66. Matsuno, H. *et al.* Antirheumatic effects of humanized anti-Fas monoclonal antibody in human rheumatoid arthritis/SCID mouse chimera. *J. Rheumatol.* **29**, 1609–1614 (2002).
67. Vita, S. D. *et al.* Efficacy of selective B cell blockade in the treatment of rheumatoid arthritis: evidence for a pathogenetic role of B cells. *Arthritis Rheum.* **46**, 2029–2033 (2002).
68. Wang, H. *et al.* TACI-ligand interactions are required for T cell activation and collagen-induced arthritis in mice. *Nature Immunol.* **2**, 632–637 (2001).
69. Banda, N. K. *et al.* Mechanisms of effects of complement inhibition in murine collagen-induced arthritis. *Arthritis Rheum.* **46**, 3065–3075 (2002).
70. Firestein, G. S. & Zvaifler, N. J. How important are T cells in chronic rheumatoid synovitis? II. T cell-independent mechanisms from beginning to end. *Arthritis Rheum.* **46**, 298–308 (2002).
71. van der Heijden, I. M. *et al.* Presence of bacterial DNA and bacterial peptidoglycans in joints of patients with rheumatoid arthritis and other arthritides. *Arthritis Rheum.* **43**, 593–598 (2000).
72. Wälscher-Kuntz, D. R., Weyand, C. M., Fulbright, J. W., Moore, S. B. & Goronzy, J. J. HLA-DRB1 molecules and antigenic experience shape the repertoire of CD4 T cells. *Hum. Immunol.* **44**, 203–209 (1995).
73. Gebe, J. A. *et al.* T cell selection and differential activation on structurally related HLA-DR4 ligands. *J. Immunol.* **167**, 3250–3256 (2001).
74. Zanelli, E., Breedveld, F. C. & de Vries, R. R. P. HLA association with autoimmune disease: a failure to protect? *Rheumatology* **39**, 1060–1066 (2000).
75. Roudier, J., Petersen, J., Rhodes, G. H., Luka, J. & Carson, D. A. Susceptibility to rheumatoid arthritis maps to a T-cell epitope shared by the HLA-Dw4 DR beta-1 chain and the Epstein-Barr virus glycoprotein gp110. *Proc. Natl Acad. Sci. USA* **86**, 5104–5108 (1989).
76. Aceves-Avila, F. J., Baez-Molgado, S., Medina, F. & Fraga, A. Paleopathology in osseous remains from the 16th century. A survey of rheumatic diseases. *J. Rheumatol.* **25**, 776–782 (1998).
77. Rothschild, B. M., Turner, K. R. & DeLuca, M. A. Symmetrical erosive peripheral polyarthritis in the Late Archaic Period of Alabama. *Science* **241**, 1498–1501 (1988).

Acknowledgements Supported by grants from the National Institute of Arthritis and Musculoskeletal and Skin Diseases. The author also thanks N. J. Zvaifler for helpful discussions on this topic as well as two decades of instruction, collaboration and friendship.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/

Iberia/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel +49 89 54 90 57 11/-2

Fax +49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707
Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),
19-1 Haraikatamachi,
Shinjuku-ku,
Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejp.com

naturejobs

Biotech's balancing act

The number of jobs within Europe's biotechnology sector fell last year for the first time in a decade. According to a report published last week by analysts Ernst & Young, the number of positions in 2002 was 82,124 — 6% fewer than the previous year. But despite the bad news, the company's report, *Endurance*, features some encouraging signs for people who are thinking about joining the industry.

The main source of the sector's woes is the stock market — or more specifically, investors' growing impatience with biotech companies that have yet to show products or profits. European biotech valuations declined by 50–60% last year. And, as a result of the weak stock market, raising new money on the market has been difficult — only €123 million (US\$140 million) was raised last year, compared with €5.5 billion in 2000.

The financial news isn't all bad — especially for the better-established companies. Venture capital has continued to flow, but mostly towards firms that are close to bringing a product to market. In addition, all of the top ten European biotech companies have product sales and eight of those companies are profitable. And between them, Europe's public biotech firms have 53 products currently undergoing phase III clinical trials.

As in most industries, jobs in the biotechnology sector tend to follow the money. A few years ago, that meant a boom for start-up companies. But firms with no products on the market — or at least in clinical trials — are no longer good bets. Those that do have products close to approval, on the other hand, should continue to attract investment, and so are likely to offer stronger employment prospects. Anyone interested in pursuing a career in Europe's biotech sector may well want to consider their chosen company's financial stability and product pipeline before making their final decision.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Molecular biology
in Heidelberg

p364

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Language lessons

Counting the benefits: Cellzome is one of the spin-offs from the European Molecular Biology Laboratory that hosts an international staff of scientists.

Heidelberg

Signs in the corridor of the European Molecular Biology Laboratory (EMBL) in Heidelberg say “Finally learn German”. Those advertisements for language lessons reflect one of several apparent contradictions that Heidelberg offers. It is a German city where German sometimes feels like a second language, a cosmopolitan centre, even though the population is less than 150,000. And EMBL, for its part, is a potential template for

European biology even though its organization is based on a Swiss physics lab.

Why should EMBL be given such consideration? Two indicators provide a hint. It is the only European laboratory to make the top 15 in the list of most-cited molecular-biology and genetics labs drawn up by ISI — the former Institute for Scientific Information — for 1992–2002. And it draws seven PhD applicants for every available place.

Those statistics have caught the eye of European Commission members, who have discussed using EMBL as a model for European science. If that is to happen, national labs in Germany and other countries will have to turn their organizations upside down: in many ways EMBL offers the antithesis of a traditional European scientific career structure. It emphasizes youth and turnover rather than experience and stability, broad curiosity over narrow specialization, and equal opportunities for scientists from outside Germany.

YOUTH CULTURE

Although EMBL is most famous for these qualities, they have been absorbed and expanded upon by many of the other Heidelberg institutions, both public and private. The Centre for Molecular Biology (ZMBH) at the University of Heidelberg also recruits young group leaders, often straight from their first postdoc. And the German Cancer Research Centre (DKFZ), despite its national base, has one scientist who is jointly funded by the DKFZ and the French biomedical research agency INSERM, but who also does some work with China.

On the industrial side, a trio of EMBL spin-offs — Cellzome, LION Bioscience and Anadys Pharmaceuticals (see ‘Spinning out’, left) — also host an international staff of scientists, many of whom passed through at least one of the city’s academic institutions early in their career. And multinational companies wanting to

Spinning out

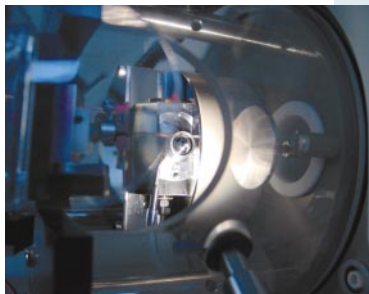
Three Heidelberg spin-off companies had similar origins but divergent employment opportunities. LION Bioscience, Anadys Pharmaceuticals and Cellzome all emerged from the European Molecular Biology Laboratory (EMBL) during the past few years, built on the foundations of bioinformatics and intending to become drug companies.

LION, the largest and oldest, turned away from that path at the end of 2002, when it closed its biology labs to focus solely on providing bioinformatics tools,

service and support. The company went from a high of 500 employees to 360 now, 200 of which are based in Heidelberg. There is a freeze on new positions until late 2004, when the company is expected to break even.

Anadys announced in March that it had completed early clinical work for a hepatitis drug and plans to take on medicinal chemists and other scientific staff. Cellzome, although it doesn’t have any drugs in the clinical pipeline yet, managed to raise E 30 million (US\$33 million) this spring and is also looking to expand — especially in terms of drug discovery. **P.S.**

The ion source of a mass spectrometer at Cellzome.



take advantage of the academic expertise have outposts nearby, including Merck, Schering-Plough and Roche Diagnostics.

This means that you are as likely to hear English being spoken in the city as you are to hear German, and more likely in EMBL's labs, corridors and cafeteria. "The reality is that you don't need to speak German to function in Heidelberg," says Iain Mattaj, scientific director of EMBL.

EMBL's own make-up attests to that. Mattaj estimates that its staff of 700 includes 75 nationalities. Because group leaders' contracts run for five years, with an option to renew for four more, the cultural mix changes frequently.

Mattaj is glad that the European Commission is noting EMBL's success, and hopes it will use this as a model for other labs, as he believes the turnover system brings innovation and prevents stagnation. "The way to increase European performance is to have a lab like EMBL," he says.

MIX AND MATCH

As for getting into EMBL as a student, Mattaj recommends not being too focused on one problem or one aspect of biology. "What we are looking for are people who are very motivated to do research and to think broadly," he explains. An ability or willingness to mix maths or physics with biology also helps, he adds.

For example, biologists working closely with mathematicians at EMBL helped to improve their use of microarrays, says Wilhelm Ansorge, programme coordinator for biochemical instrumentation. He adds that EMBL's multidisciplinary turnover model was based on that of his previous employer CERN, the European laboratory for particle physics near Geneva. "The important thing is that everyone has to learn a bit about the fields of the others," he says.

That sort of cooperation is not limited to EMBL, says Renato Paro, director of the ZMBH. For example, microarray techniques developed through a collaboration between the DKFZ and the ZMBH were later passed on to scientists at EMBL working on a different project. "In that sense, it's not competitive," Paro says of the interaction between the Heidelberg academic research institutions.

In some ways, the ZMBH is structured similarly to EMBL. It employs young group leaders on a competitive basis. Although the turnover is not built in to the extent it is at EMBL, there is still plenty of mobility, with several group leaders in recent years being recruited as directors of Max Planck institutes.

One of the biggest differences between the two organizations is that ZMBH group leaders have to teach, which cuts into their research time. But as a result, they are treated more like senior lecturers, with the same vote in academic affairs.

And like EMBL, the ZMBH provides scientific colloquia and seminars in English that are open to anyone. Last year, the two banded together with the DKFZ to sponsor the Heidelberg Forum on the Biosciences and Society, which aims to inform the public about current issues in science.

Although such meetings are good for getting people to talk, Paro would like to see the Heidelberg research institutes sponsor more informal events, where only the food and beverages are planned and

there are no formal scientific presentations. Such activities, he says, would mimic the give-and-take that occurs daily in EMBL's cafeteria and lounges. "EMBL has recognized the importance of exchanging information in a less formal environment. Why not expand that and try it locally?" asks Paro.

With the mix of backgrounds at Heidelberg, such conversations can be linguistically interesting, says Jean Rommelaere, a DKFZ cancer researcher. His mixed background provides a testament to that. The Belgian scientist is jointly funded by the DKFZ and INSERM.

Rommelaere, in his turn, estimates that in his lab's ten years, he has had members from 17 different countries — several from China, where he is fostering relationships by, among other things, hosting students and postdocs in his labs.

"We start a session in French, continue in German, and when we don't understand each other, finish in English," says Rommelaere. It all works out in the end, because in Heidelberg, science is the common language.

Paul Smaglik is editor of *Naturejobs*.



Jean Rommelaere has postdocs and students from 17 different countries.

The Munster match

A fledgling joint research programme in clinical medicine between the European Molecular Biology Laboratory (EMBL) and the University of Heidelberg Medical School was made possible when two MDs who received their training in Munster, Germany, more than 20 years ago were last year reunited in Heidelberg.

Matthias Hentze, the dean of graduate studies at EMBL, had retained his research ties with Andreas Kulozik since their medical-school days. So when Kulozik moved from Berlin to Heidelberg last year, it made good sense to start up a more institutional collaboration. The two make a good team — "a match made in Munster", Hentze says — because EMBL can provide basic

biology and mouse models, and the medical campus offers clinical facilities.

The Molecular Medicine Partnership Unit lets students do 1–3 years of basic research along with conventional medical training. The programme has two junior residents and two postdocs, but Hentze and Kulozik expect to have 25 by the end of 2004 — especially if they set up a formal MD/PhD programme.

P.S. www.embl-heidelberg.de

Team spirit: Matthias Hentze (left) and Andreas Kulozik.

